GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:26:08 ; Search time 41.26 Seconds
(without alignments)
3391.977 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	938	14.4	1537	15	AAR60562	Yeast 4.7 kB agglu
2	721.5	11.1	5179	22	AAM24516	C899P predicted am
3	685	10.5	2586	22	ABB66878	Drosophila melanog
4	640.5	9.9	2344	22	AAU37120	Staphylococcus aur
5	601	9.3	1795	22	ABB69806	Drosophila melanog
6	553.5	8.5	2870	21	AAY95559	Caenorhabditis ele
7	553.5	8.5	3178	21	AAY95556	Caenorhabditis ele
8	545.5	8.4	957	21	AAY59288	Human MUC11 polype
9	545.5	8.4	957	22	AAM24513	C900P predicted am
10	534	8.2	849	17	AAW06725	FLO1 protein, invo
11	516	7.9	894	15	AAR47578	Flocculation prote
12	516	7.9	894	15	AAR58754	S. cerevisiae FLO1
13	507	7.8	1532	21	AAB40945	Human ORFX ORF709
14	488	7.5	1721	19	AAW48299	Cryptosporidium pa
15	488	7.5	1721	21	AAB11727	Portion of Cryptos
16	487	7.5	1837	21	AAB11726	Cryptosporidium pa
17	468.5	7.2	862	15	AAR60563	Yeast 2.6 kB agglu
18	463	7.1	688	22	ABB30137	Peptide #2788 enco
19	463	7.1	688	22	ABB35307	Peptide #2813 enco
20	463	7.1	688	22	ABB20749	Protein #2748 enco
21	463	7.1	688	22	AAM56138	Human brain expres
22	463	7.1	688	22	AAM68511	Human bone marrow
23	463	7.1	688	22	AAM16315	Peptide #2749 enco
24	463	7.1	688	22	AAM28810	Peptide #2847 enco
25	463	7.1	688	22	AAM04053	Peptide #2735 enco
26	450.5	6.9	4498	22	ABB58595	Drosophila melanog
27	435.5	6.7	1045	22	ABG08332	Novel human diagno
28	428	6.6	1237	21	AAY81609	Streptococcus pneu
29	417.5	6.4	746	22	ABB59201	Drosophila melanog
30	417	6.4	560	22	AAU37464	Staphylococcus aur
31	404.5	6.2	2035	15	AAR57141	Host cell factor p
32	404	6.2	2570	22	ABG06375	Novel human diagno
33	402	6.2	2112	22	ABB60403	Drosophila melanog
34	392	6.0	502	22	AAU34408	Staphylococcus aur
35	392	6.0	1638	20	AAY00138	Enterococcus faeca
36	392	6.0	1638	20	AAY00140	Enterococcus faeca
37	392	6.0	1638	20	AAY00142	Enterococcus faeca
38	390	6.0	596	21	AAY99408	Human PRO1342 (UNQ
39	390	6.0	596	22	AAU29178	Human PRO polypept
40	390	6.0	596	22	AAB87575	Human PRO1342. Ho
41	390	6.0	596	22	AAB66157	Protein of the inv
42	379	5.8	1872	22	ABG15263	Novel human diagno
43	378.5	5.8	1976	22	ABB69419	Drosophila melanog
44	377	5.8	2858	22	ABB71150	Drosophila melanog
45	377	5.8	3060	22	ABB58064	Drosophila melanog

ALIGNMENTS

RESULT 1
AAR60562
ID AAR60562 standard; Protein; 1537 AA.
XX
AC AAR60562;
XX
DT 13-APR-1995 (first entry)
XX
DE Yeast 4.7 kB agglutination gene FLO1L.
XX
KW Yeast; agglutination; FLO1L.
XX
OS Saccharomyces cervisiae ABXL-1D.
XX
PN WO9419475-A.
XX
PD 01-SEP-1994.
XX
PF 24-FEB-1994; 94WO-JP00290.
XX
PR 26-FEB-1993; 93JP-0038871.
XX
PA (PANI-) PANIMOLABORATORIO BRYGGERILABORATORIUM.
PA (SAPB) SAPPORO BREWERIES.
XX
PI Keraenen S, Ogawa M, Onnela M, Penttila M, Takata Y;
PI Watari J;
XX
DR WPI; 1994-294338/36.
DR N-PSDB; AAQ71390.
XX
PT New yeast agglutination genes and yeast contg. them - impart
PT agglutination properties to facilitate removal from fermentation
PT media

XX
PS Disclosure; Page 43-48; 75pp; English.
XX
CC The agglutination gene is called FLO1. Saccharomyces cerevisiae
CC includes an agglutination gene of 4.7 kb (FLO1L) and an
CC agglutination gene of 2.6 kb (FLO1S). FLO1L is the intact FLO1 gene
CC on chromosome I, and FLO1S is the FLO1L gene with a portion of the
CC ORF deleted in frame. FLO1L imparts a relatively strong
CC agglutinative property to the host yeast into which it is introduced,
CC while FLO1S imparts a weaker agglutinative property.
XX
SQ Sequence 1537 AA;

Query Match 14.4%; Score 938; DB 15; Length 1537;
Best Local Similarity 26.0%; Pred. No. 3e-40;
Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;

```
Qy       2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
Db      10 lavftllal-tsvasgateaclpagqrksgmninfyqyslkdsstysnaaymaygyaskt 68

Qy      41 ----------------------------KGPGYPTW-----------NAVLGWSLD--G 58
Db      69 klgsvgggtdisidynipcvsssgtfpcpqedsygnwgckgmgacsnsqgiaywstdlfg 128

Qy      59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
Db     129 fyttptnv-tlemtgyflppqtgsytfkfatvddsailsvggatafnccaqqqppitstn 187

Qy      86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
Db     188 ftidgikpwggslppniegtvymyagyyypm---------kvvysnavswgtlpisvtlp 238

Qy     132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
Db     239 -----dgttvsddfeg--------yvysfdd-------dlsqsnctvpdpsnya-vsttt 277

Qy     191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
Db     278 tttepwtgtftststemttvtgtngvptdetvivirtpttastiitttepwnstftstst 337

Qy     251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
Db     338 elttvtgtngvrtdetii------virtpttattaitttepwnstftstst-------- 384

Qy     311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT----------- 355
Db     385 ------ttvtgtnglptdetiivirtpttattamtttqpwndtftststemttvtgtngl 438

Qy     356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
Db     439 ptdetiivirtpttattamtttqpwddtftststemttvtgtnglptdet-iivirtptt 497

Qy     399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
Db     498 attamtttqpwndtftststemttvtgtnglptde-tiivirtpttattamtttqpwndt 556

Qy     448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
Db     557 ftststemttvtgtnglptdetiivirtpttattaitttepwnstftststelttvtgtn 616

Qy     495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
Db     617 glptdetiivirtpttattamtttqpwndtftststemttvtgtnglptdetiivirtpt 676

Qy     543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
Db     677 tattamtttqpwndtftststemttvtgtnglptdetiivirtpttattamtttqpwndt 736

Qy     592 YATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTVT--- 635
Db     737 ftststemttvtgtngvptdetvivirtptseglistttepwtgtftststemttvtgtn 796

Qy     636 GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
Db     797 gqptdetvivirtptseglvtttepwtgtftststemttitgtngvptdetvivirtpt 856

Qy     687 NHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEYWSQS 735
Db     857 seglistttepwtgtftststemttitgtngqptdetvivirtptseglisttttepwtgt 916

Qy     736 YATT----TTVTAPPG-GTD--TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPPGGT- 785
Db     917 ftststemttvtgtngvptdetvivirtptseglistttepwtgtftststevttitgtn 976

Qy     786 -----DTVIIYESMSSSKI------------STSSNDITSIIPSFSRP------------ 816
Db     977 gqptdetvivirtptseglistttepwtgtftstsaemttvtgtngqptdetvivirtpt 1036

Qy     817 --HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICSDGKE 869
Db    1037 seglvttttepwtgtftststemstvtgtnglptdetvivvktpttaisssslssssgqi 1096

Qy     870 CSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
Db    1097 tssitssrpiit-pfypsngtsvisssvisssvtsslftsspvisssvisssttttstsif 1155

Qy     928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
Db    1156 sessksv--iptssstsgssesetssagsvsss----sfissessksptyssss----l 1205

Qy     988 EIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITEDITT 1035
Db    1206 plvtsattsqetasslppatt---------tktseqttlvtvtsceshvctesispaivs 1256

Qy    1036 SQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
Db    1257 tatvtvsgvtteyttwcpisttettkqtkgtteqttettkqttvvtisscesdvcsktas 1316

Qy    1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELVATT- 1145
Db    1317 paivststatingvtteyttwcpist--tesrqqttlvtvtscesgvcsetaspaivsta 1374

Qy    1146 ----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSVTGGA 1188
Db    1375 tatvndvvtvyptwrpqtaneesvsskmnsatgetttntlaaetttntvaaetitntgaa 1434

Qy    1189 -------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
Db    1435 etktvvtsslsrsnhaetqtasatdvighsssvvsvsetgntksltssglstmsqqprst 1494

Qy    1225 SP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db    1495 passmvgystasleistyagsansllagsglsvfiaslllai 1536
```

RESULT 2
AAM24516
ID AAM24516 standard; Protein; 5179 AA.
XX
AC AAM24516;
XX
DT 12-OCT-2001 (first entry)
XX
DE C899P predicted amino acid sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX

PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 446-462; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 5179 AA;

Query Match 11.1%; Score 721.5; DB 22; Length 5179;
Best Local Similarity 25.3%; Pred. No. 2.1e-28;
Matches 335; Conservative 158; Mismatches 554; Indels 279; Gaps 54;

```
Qy      59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118
           |:  |  | | : |     ||:  |  :|       |       |  | :| | :
Db    1647 ttttpppttpspp-----ttttpsppittttppptt--------tpsspitttpsppt 1693

Qy     119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
             :      | |         | |||        |  | : |        ::      :|:
Db    1694 ttmttpspttpsspittttttpsstttpsppptttmttpspttpspptttm------ttl 1747

Qy     179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224
           |   |: |  : : ||:   |         |    | |   |:  ||   |    ||
Db    1748 pptttsspitttplppsitpptfspfstttpttpcvplcnwtgwldsgkpnfhkpggdte 1807

Qy     225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272
           :      ||   |   |   :|  :  |         |  | | |:  | ::  | |
Db    1808 l------igdvcg-pgwaaniscratmypdvpigqlgqtvvcdvsvglicknedqkpggv 1860

Qy     273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332
             |           : | |  :  |      ::| |:  |  :|    :   :   ||
Db    1861 ipm----------afclnyeinvqcc--ecvtqpttmtttttenptpptttpitttttv 1907

Qy     333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIG-ETATV 391
           | | : |   |   |   |      |  | | ||: :  ||:      | | :| |
Db    1908 tptptptgtqtptttpitttttvtptptptgtqtptttpitttttvtptptptgtqtptt 1967

Qy     392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432
               |  ||||||   | |         |||||          |:|  |  | |     |
Db    1968 ---tpitttttvtptptptgtqtptttpitttttvtptptptgtqtptttpitttttvtp 2024

Qy     433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492
           | || : |   :    ||||||  |  | |   : |    :|||        || | |
Db    2025 tptptgtqtp-tttpitttttvtptptptgt---qtptttpittt--------ttvtptp 2072

Qy     493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535
           | ||        :      | || | |:  : :  ||||   |   | ||        :
Db    2073 tptgtqtptttpitttttvtptptptgtqtptttpitttttvtptptptgtqtptttpit 2132

Qy     536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592
               | || | |:  : :  ||||   |   | || :     |    :|||
Db    2133 ttttvtptptptgtqtptttpitttttvtptptptgtqt-----ptttpittt------- 2180

Qy     593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643
           || | |  | || |     |     :|||         || | |  |:||          |
Db    2181 -ttvtptptptgtqt-----ptttpittt--------ttvtptptptgtqtptttpittt 2226

Qy     644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703
             :   | || | |   :    ||||:|  |  | |          | |||        ||
Db    2227 ttvtptptptgtqtp-tttpitttttvtptptptgt--------qtptttt-----pittt 2272

Qy     704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754
           ||||  |  | |   : |    :|||        || | |  | ||          |  :
Db    2273 ttvtptptptgt---qtptttpittt--------ttvtptptptgtqtptttpitttttv 2321

Qy     755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814
             | || | |   :    ||||||  |  | |    :: ::: |:|:    |:: |: :
Db    2322 tptptptgtqtp-tttpitttttvtptptptgt----qtptttpittt----ttvtptpt 2372

Qy     815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874
               ||: ::|   ::  |||  :  |    :|| :| : | |       | :
Db    2373 ptgtqtptttpitt---tttvtptp-tptgtqtptttpitttttvtptptptgtq----- 2423

Qy     875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932
               | :   :  |:  || :     :|  |: |:  :   : :|     | | :
Db    2424 -------tpttttpitttttvtptptptgtqtptttpitttttvtptptptgtqtptttpi 2476

Qy     933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980
           ||  |         |  | :: :|: | ::  |: : :::  |    |:     : |
Db    2477 tttttvtptptptgtqtptttpitttttvtptptptgtqtptttpitttttvtptptptg 2536

Qy     981 QSTSTEIEIVTTSSTKVLP-----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028
             | |    | ||::    |           |: :: |   |  || |:   ||  ||: :
Db    2537 tqtptttpitttttvtptptptgtqtptttpitttttvtptptptgtqtptttpitttt 2596

Qy    1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087
           :|    |  ||   |  | :| |: |  | |   : :    :  :  : :|:: |:
Db    2597 vt---ptptpt---gtqtptttpitttttvtptptptgtqtptttpitttttvtptptpt 2650

Qy    1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ----------- 1127
              | :|    |||: |  |:|          ::: | ||  || | |
Db    2651 gtqtptttpitttttvtptptptgtqtptttpitttttvtptptptgtqtptttpitttt 2710

Qy    1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
                ||   | : | : :  ||  |       | ::|:|   :: || |    |   ::
Db    2711 tvtptptptgtqtptttpitttttvtptptptgtqtptttpitttttvtptptptgtqtp 2770

Qy    1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
           |:  :|       : :  |: :|:   : : |: ||  | :|  |::  :| | :  |
Db    2771 tttpitttttvtptptptgtqtptttpitttttvtptptptgtqtptttpitttttvtpt 2830

Qy    1236 YDGSGS 1241
              :|:
Db    2831 ptptgt 2836
```

```
RESULT   3
ABB66878
ID   ABB66878 standard; Protein; 2586 AA.
XX
AC   ABB66878;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 27426.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10981.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 27426; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.

     Sequence   2586 AA;

  Query Match             10.5%;  Score 685;  DB 22;  Length 2586;
  Best Local Similarity   24.7%;  Pred. No. 6.8e-27;
  Matches  405;  Conservative  183;  Mismatches  557;  Indels  494;  Gaps   67;

Qy     15 ASAKTITGVF-DSFNSLTWSNAANYAFKGPGYPTWNAVLGWSL--DGTSANPGDTFTLNM 71
Db    102 astntitkiiddglssqttsssa---------pvvdvtqgsssngdgnstqsstttt--- 149

Qy     72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL---KSSIKAFGTV 128
Db    150 ------tttttssd--------------ggefttssdpvvevsqgtnggnsstqsssst 188

Qy    129 TLPIAFNVGGTGSSTD---------------LEDSKCFTAGTNTVTFNDGDKDISID-- 170
Db    189 ttttssdegqttsssdpvvevaqgsssngdgnstqsltttttttttssdggqsttssdpv 248

Qy    171 VEFEKSTVDPSAYLYAS----RVMPSLNKVTTLFVAPQCE-------NG--------YTS 211
Db    249 vevsqgtnggnsstqsssstttttssdegqttsssdpvvevaqgsssngdgnstqsstt 308

Qy    212 GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG 271
Db    309 ttttttssdgggqsttlsdpvvevsqgtnggnsstqsssstttttssdeg-qttsssdpvg 367

Qy    272 YRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAG-----SNGIV 326
Db    368 ----------------evaqgsssngdgnstqsst-tttttttssdggqsttssdpvv 409

Qy    327 IVA------TTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVG------- 373
Db    410 evsqgtnggnsstqsssattttt-----ssdegqttsssdpvsevaqgsssigdgnstqs 464

Qy    374 -VTTSYLTKTAPIGETATV----IVDVPYHTTTTVTSEWTGTITTTTTRTN---PTDSID 425
Db    465 sttttttttssdgggqsttssdpvvevsqgtnggnsstqsssstttttssdegqttsssd 524

Qy    426 TV--VVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGG-----TDTVI-----------IR 467
Db    525 pvvevaqgsssngdgnstq--sstttttttttssdgggqsttssdpvveasqgtnggnsst 582

Qy    468 EPPNHTVTTT---------------------------EYWSQSFATTTTVTAPPGG--- 496
Db    583 qsssstttttssdegqttsssdpvsevaqgsssigdgnstqsstttttttttssdggqst 642

Qy    497 TDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWS--- 553
Db    643 tssdpvveasqgtnggnsstqsssstttttssdegqtts-----ssdpvvevaqgsssng 697

Qy    554 ------QSYATTTTVTAPPGGTDSVIIREP-----------------PNHTVTTT----- 585
Db    698 dgnstqsstttttttttssdgggqsttlsdpvvegsqgtnggnsstqsssstttttssdeg 757

Qy    586 ----------------------EYWSQSYATTTTVTAPPGGTDT------VIIREPPNH 616
Db    758 pttsssdpvvevaqgsssngdgnstqsstttttttttssdggesttssdpvvevsqgtng 817

Qy    617 TVTTTEYWSQSFATTTT----------------VTGPPSGTDTVIIREPPNPTVTTTEY 659
Db    818 dnsstq--sssstttttssdegqttsssdpvsevaqgsssstgdgnstqsstttttttttdi 875

Qy    660 W---SQSYATTTTITAPPGETDT-----VLIREPPNHTVTTTEYWSQSYATTTTVTAPPG 711
Db    876 frcprqrtttttttssdggesttlsdpvvevsqgtnggnsstq---sssstttttssdeg 932

Qy    712 ETDT-----VLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDT------VIIREPPNP 760
Db    933 qttsssdpvvevaqgsssngdgnstqsstttttttttssdggesttssdpvvevsqgtng 992

Qy    761 TVTTTEYWSQSFATTTTVTAPPGGTDT-----VIIYESMSSSK---ISTSSNDITSIIPS 812
Db    993 dnsst----qsssstttttssdegqatsssapvvdisqgsssngdgnstqsstttttittt 1048

Qy    813 FSRPHYVNSTTSD-----------LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTT 860
Db   1049 tssdgdqsttssdpvvevsqgtnggnsstqsssstttttssdegqttssfgscrrshsri 1108

Qy    861 ELIC---------------------------SDGK--ECSRLSSSSGIVTNPDSNESSI 890
Db   1109 lfqcprqrqqlqrrhlpmvanppphltpvvevsqgtngdnsstqsssstttttssdegqt 1168

Qy    891 VTSTVPTASTMSDSLSST--DG-ISATSSDNV-------------SKSGVSVTTETSVTT 934
Db   1169 tsssapvsstttttttttssdgdqsttssdpvvevsqgtngdnsstqsssstttttssde 1228

Qy    935 IQTTP-----------------NPLSSSVTSLTQLSSIPSVSESES------KVTFTSN 970
Db   1229 gqttsssapvvevtqgsssngdgnstqsstttttttttssdggesttssdpvvevsqgtn 1288

Qy    971 GDNQSGTHDSQSTSTEIEIVTTSS-----TKVLPPVV------SSNTDLTSEPTNTREQP 1019
Db   1289 gdn-sstqsssstttt-----ttssdegqttsssapvvevtqgsssngdgnstqsst--tt 1340

Qy   1020 TTLSTTSNSITEDITTSQPT-----GDNGDNTSSTNPVPTVATSTLASASEEDNKSG--- 1071
```

```
Db   1341 ttttttssdggesttssdpvvevsqgtngdn-sstqsssstttttssdegqttsssgpvv 1399

Qy   1072 --SHESAS-----------------------TSLKPSMGENSGLT------------TS 1093
Db   1400 disqgsssnvldndnnynddifrwwanpppyltplwksvkeqmaitaplslrrppqqqts 1459

Qy   1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE----------LVA 1143
Db   1460 sd-egqtts-ssapvvdisqgsssngdgnst-qsstttttttttssdggesttlsdpvve 1516

Qy   1144 TTQATNENGGKSPSTDLTSSLTTGTSA----STSANSELV--------------TSGSVT 1185
Db   1517 vsqgtn---gdnnstqsssstttttssdegqttsssdpvvevaqgsssngdgnstqsstt 1573

Qy   1186 GGAVASASNDQSHST----------------SVTNSNSIVSNTPQTTLSQQVTSSSPST 1228
Db   1574 ttttttssdggesttssdpvvevsqgtngdnsstqsssstttttssdegqttsssdpvv 1633

Qy   1229 NTFIASTYDGSGSIIQHST 1247
Db   1634 evaqgsssngdgnstqsst 1652
```

RESULT 4
AAU37120
ID AAU37120 standard; Protein; 2344 AA.
XX
AC AAU37120;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1290.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS54979.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 12713; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2344 AA;

Query Match 9.9%; Score 640.5; DB 22; Length 2344;
Best Local Similarity 22.2%; Pred. No. 1.2e-24;
Matches 323; Conservative 269; Mismatches 542; Indels 323; Gaps 46;

```
Qy     17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76
Db    297 aitlgtridsnksfhfsgkvnlgnkyeghgnggdgigfaf-----spg------------ 339

Qy     77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIK--------AFGTV 128
Db    340 -vlgetglngaavgigglsnaf--gfkldtyhn-tstpnssakakadpsnvagggafgaf 395

Qy    129 TLPIAFNVGGTGSSTDLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186
Db    396 vttdsygvastytssstadnaaklnvqptnn-tfqd------fdinyngdt--------- 439

Qy    187 SRVMPSLNKVTTLFVAPQ----------CENGYTSGTMGFSSSNGDVAIDCSNIHIGITK 236
Db    440 --------kvmtvkyagqtwtrnisdwiaksgttnfslsmtastgg-atnlqqvqfg--- 487

Qy    237 GLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG--------YRPFIDAYISATDVNQY 288
Db    488 -----------tfeytesavt---qvryvdvttgkdiippktysgnvdqvvtidnqqsa 532

Qy    289 TLAYTNDYTCAGSRLQS--------------KPFTLRWT---------GYKNSDAGS-- 322
Db    533 ltakgynytsvdssyastyndtnktvkmtnagqsvtyyftdvkaptvtvgnqtievgktm 592

Qy    323 NGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPI------------------PT 364
Db    593 npivltttdngtgtvtntvtglpsglsydsatnsiigtptkigqstvtvvstdqannkst 652

Qy    365 TTITTSYVGVTTSYLTKTAPIGETATVI------------------------------ 392
Db    653 ttftinvvdttaptvt---pigdkssevfspispiniatqdnsgnavtntvtglpsgltf 709

Qy    393 -----------VDVPYHTTTTVTSEWTGTITTTT-----TRTNPTDSIDTVVVQVPLPNP 436
Db    710 dstnntisgtptnigtstitivstdasgnkttttfkyevtrnsmsdsvst--sgstqqsq 767

Qy    437 TVSTTEYWSQSFATT------TTVTAPPGGTDTVIIRE--PPNHTVTTTEYWSQSFATTT 488
Db    768 svstskadsqsaststsgsimtstsastskstsvslsdsvsaskslstsesnsvssstst 827

Qy    489 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT 548
Db    828 slvnsqsvsss--msgsvskstslsdfisnsssteksesvststsdslrtstslsdsvsm 885

Qy    549 TEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAP------- 601
Db    886 stsgslsksqslststsdsastsqsvsdstsnsistseslsesgstsesisisnsisnsv 945

Qy    602 -------PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTV 654
Db    946 sastsklesqstsislstsdsksmstseslsdststsdsvsgslsvagsqsvststsdsm 1005

Qy    655 TTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT-------TTVT 707
```

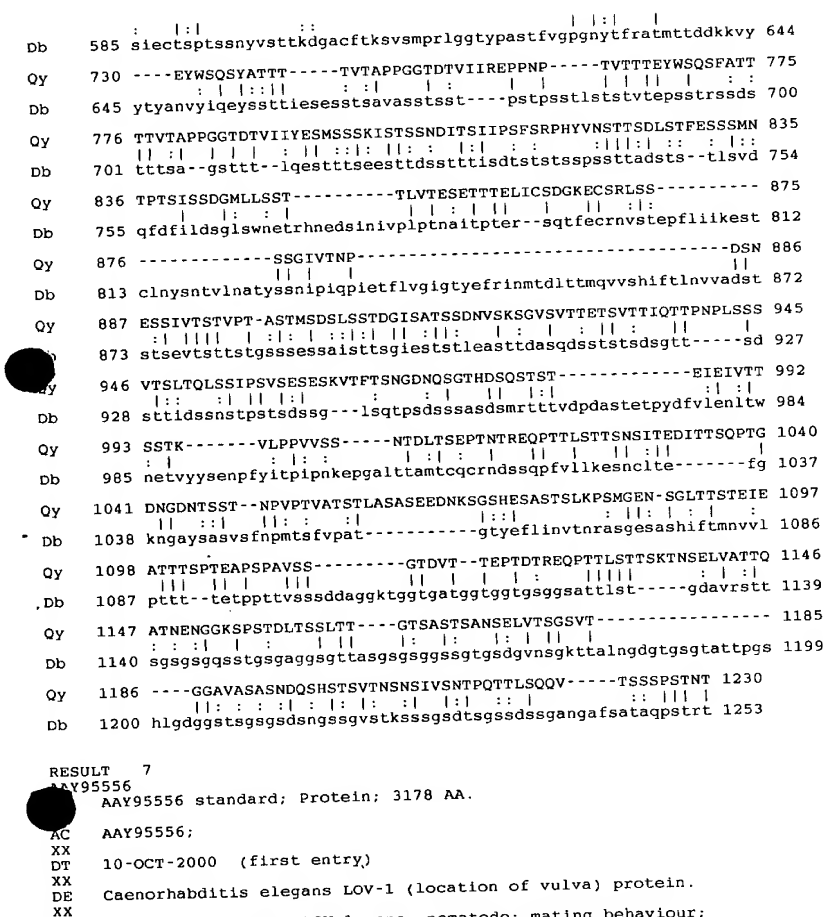
```
Db    1006 stsemisdsmstsgslaasdsksmsvsssmstsqsgstseslsdsistsdsdskslslst 1065

Qy     708 APPGETDTVL-----IREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTV 762

Db    1066 sqsgststststsssvrmsesqstsgsmstsqsdstsistsfsdstsds------ksast 1119

Qy     763 TTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNST 822

Db    1120 assesisqsvststs-----gsvststslstsnsertstsmsdstslstsesdstsdsts 1174

Qy     823 TSDLSTFESSSMNTPTSIS-------SDGMLLSSTTLVTE------SETTTELICSDGKE 869

Db    1175 tsd-siseaisgsestsislsesnstsdsesksasaflseslsestsestseslsgstsd 1233

Qy     870 CSRLS-------SSSGIVTNPDSNESSIVTSTVPTAST---------------MSDSLSS 907

Db    1234 stslsdsnsesgststslsnstsgsasiststsgsaststvksesvstslststststslsd 1293

Qy     908 TDGISATSSDNVSKS-----GVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESE 962

Db    1294 stslstslsdstsgsksnslsasmstsdsistrkseslsastslsgstsesesgstssse 1353

Qy     963 SKVTFTS---------NGDNQSGTHD--SQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE 1011

Db    1354 sksdstsmslsmsqstsgstsvstseslsdststslslsasmnqsgvdsnsasqsastst 1413

Qy    1012 PTNTRE--QPTTLSTTSNSITEDITTSQPT--GDNGDNTSSTNPVPTVATSTLASASEED 1067

Db    1414 ststsesdsqstssytsqstsqseststststslsdstsiskstsqsgststsaslsgsese 1473

Qy    1068 NKSGS---------HESASTSLKPSMG-ENSGLTTSTEIEATTTSPTEAPSPAVSSGTDV 1117

Db    1474 sdsqsistststseskseststslsdststsnsgsaststllsnsasasesdssstsl-sds 1532

Qy    1118 TTEPTDTREQPTTLSTTSKTNSELVAT-----------TQATNENGGKSPSTDLTSSLT 1165

Db    1533 tsasmqssesdsqststslsnsqststsirmstiasesvsestsesgstsestsesdsts 1592

Qy    1166 TGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSS 1225

Db    1593 tslsdsqsts----rstsasgsastststsdsrstsaststsmrt----stldsqsmsls 1644

Qy    1226 PSTNTFIASTYDGSGSI 1242

Db    1645 tststsvsdstslsdsv 1661


RESULT   5
ABB69806
     ABB69806 standard; Protein; 1795 AA.

AC   ABB69806;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 36210.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13909.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1795 AA;


  Query Match             9.3%;  Score 601;  DB 22;  Length 1795;
  Best Local Similarity   24.7%;  Pred. No. 9.4e-23;
  Matches  323;  Conservative  162;  Mismatches  499;  Indels  322;  Gaps   57;

Qy     126 GTVTLPIAFNV--GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 183

Db     182 gtifspverkclpgdqcpsteisdsgsyipgncelkfpecaeegtf-----rsptdcaly 236

Qy     184 LYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNY 243

Db     237 -ytcrlqesgtylqtrfkcp----gsnsfdlerklcrprsevdcfdfvpg---------- 281

Qy     244 PVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR- 302

Db     282 pv--------------------qvpyapqpyyppypaappl------yeeddydtgare 314

Qy     303 ----LQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVD------- 351

Db     315 qqpalkseklqvaaegfekp---slnvvvlqt--ttlepstayhkypaypsypsyeyssh 369

Qy     352 --------------------KTKTIE--ILQP-IPTTTITT--------SYVGVTTSYLT 380

Db     370 hrgkeraaenlelekegvprklklseniviqpetpataattreplndinkyqykrytygt 429

Qy     381 KTAPIGETATV------------IVDVPYHTTTTVTSEWTGTITTTTTR---TNPTDSID 425

Db     430 dkndvteapeiksplkglhlsenivilpettttttttttkpvvltcptisppdttpkpstt 489

Qy     426 TVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTD-TVIIREPPNHTVTTTEYWSQSF 484

Db     490 tavtk---stpkissteqhstttaktttttkrpttvtektssatekprttvvttt--tqkr 544

Qy     485 ATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP 544

Db     545 sttthntsp--dtkttirsttlspkttttp------stttpstttpstt---------tp 587

Qy     545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTT--TVTAPP 602

Db     588 stttp-------stttpstttpsttttvkv---sthrprttsqktttastttkktttspk 637

Qy     603 GGTDTVIIREPPNHTVTTTE---YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEY 659

Db     638 ttkttdiptsttsklstttqktttttthkftaattstekpktt-----tektstvstttkk 692

Qy     660 WSQSYATTTTITAPPGETDTVLIREPPNHT--VTTTEYWSQS---YATTTTVTAPPGETD 714
```

```
Db    693 stesspkptsstgkptttpkpstrttptttkvttttqittttplrsstettstqpptt- 751

Qy    715 TVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFAT 774
Db    752 ------tpqpttttttltvtpktsttttttekp-------itsspkpt-tttqkttstapn 797

Qy    775 TTTV-------TAPPGGTDTVIIYESMSSSKISTSSND--ITSIIPSFSRPHYVNSTTSD 825
Db    798 ttkvaittqkettptqststtiftrktttnnpeptstekpitsttp---kp---stttpk 851

Qy    826 LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDS 885
Db    852 tstvasst--ekttissp---kpttekstenpttnsv------ktsaltsstqratstts 900

Qy    886 NESSI---VTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTE-----TSVTTI-- 935
Db    901 eptkttqnittttpkpttlktstqe----attstqkvst--vtittkkatesspltt1st 954

Qy    936 ---QTTPNPL------SSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST--- 983
Db    955 eepnttpkplrtttptttsvtattritt-ttisesstettstqkpksttptstttttpkv 1013

Qy    984 -----STEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT--TLSTTSNSITEDITTS 1036
Db   1014 ttvivstqnpttttsktstvtittpnpspstqrpttttrqptsitasttsigttriptttt 1073

Qy   1037 QP----------------------TGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHE 1074
Db   1074 npqnstsstdlttvtrppcpdpdstsdkntntactqelqqv--nllelqspqkqeqftht 1131

Qy   1075 SASTSLKPSMGENSG---------LTTSTEIEATTTSPTEAPSPAVSSGTD--------V 1117
Db   1132 rthtaltgsrntlggqevpdymddapssaeaesgqattakaptmstlaaahllqklfhii 1191

Qy   1118 TTEPTDTREQPTTLSTTSKTNSEL-------------VATTQ--------------ATNE 1150
Db   1192 sttppsrehaptqrpssqpsssqrsrgvtiaqmarhnlatskpfiahslrlsiqqlastq 1251

Qy   1151 NGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVS 1210
Db   1252 krsippktlvthnttkepedseyydse--tseqytdednevldktqpramssttvaavlp 1309

Qy   1211 NTPQTTLSQ--QVTSSSPSTNTFIAST-------------YDGSGS 1241
Db   1310 avpsttterepqktssspsptkatsstttqpietttgdleydssgs 1355
```

RESULT 6
AAY95559
ID AAY95559 standard; Protein; 2870 AA.
XX
AC AAY95559;
XX
DT 10-OCT-2000 (first entry)
XX
DE Caenorhabditis elegans LOV-1 deletion allele lov-1(sy582del).
XX
KW Location of vulva; LOV-1 gene; nematode; mating behaviour;
KW polycystin; polycystic kidney disease; animal model;
KW signal transduction; mutant; mutein.
XX
OS Caenorhabditis elegans.
OS Synthetic.
XX
PN WO200040711-A2.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00521.
XX
PR 06-JAN-1999; 99US-0115127.
XX
PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Sternberg PW, Barr MM;
XX
DR WPI; 2000-452537/39.
XX
PT Isolated nucleic acid molecules from Caenorhabditis elegans useful for
PT producing transgenic nematodes with altered mating behavior for
PT identifying genes or regulatory factors involved in polycystic kidney
PT disease -
XX
PS Example 1; Page 131-139; 142pp; English.
XX
CC The present sequence is that of a deletion allele, termed
CC lov-1(sy582del), of the Caenorhabditis elegans LOV-1 protein
CC (see AAY95556), generated by genomic deletion of the PKD/channel
CC domain of the lov-1 gene by PCR. LOV-1 is an orthologue of
CC the human polycystin-1 (PKD1) gene that is defective in human
CC autosomal dominant polycystic kidney disease (PKD). The LOV-1
CC protein is involved in chemosensory or mechanosensory signal
CC transduction in sensory neurons. It is required for 2 male
CC sensory behaviours, 'response' of males to hermaphrodites, and
CC 'location' of the vulva. Nematodes such as C. elegans that
CC express mutant or wild-type LOV-1 or PKD-2 (see AAY95557) can be used
CC to study the functions of the proteins encoded by these genes, to
CC screen for other genes involved in PKD, to identify mutations
CC involved in the disease, and to screen for drugs that affect PKD.
CC Behaviours controlled by the action of the genes or gene products
CC are identified and used in the assays. Hence, an animal model is
CC provided that permits study of the etiology of PKD and provides a
CC tool to identify the genes involved in the disease pathway, to
CC identify compounds that may be used to treat or alter the disease
CC progression, lessen its severity or ameliorate symptoms.
XX
SQ Sequence 2870 AA;

Query Match 8.5%; Score 553.5; DB 21; Length 2870;
Best Local Similarity 24.3%; Pred. No. 4.8e-20;
Matches 290; Conservative 154; Mismatches 401; Indels 349; Gaps 51;

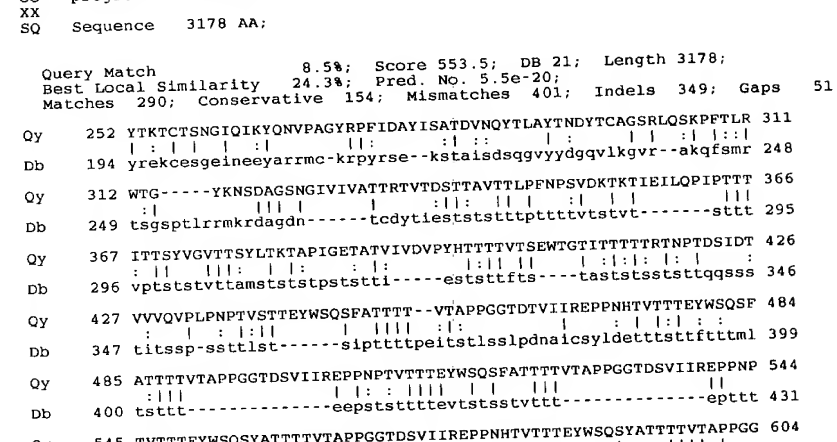
```
Qy    252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
Db    194 yrekcesgeineeyarrmc-krpyrse--kstaisdsqgvyydgqvlkgvr--akqfsmr 248

Qy    312 WTG-----YKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 366
Db    249 tsgsptlrrmkrdagdn------tcdytiestststttptttvtstvt-------sttt 295

Qy    367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
Db    296 vptststvttamstststpststti-----eststtfts----tastststsstqqsss 346

Qy    427 VVVQVPLPNPTVSTTEYWSQSFATTTT--VTAPPGGTDTVIIREPPNHTVTTTEYWSQSF 484
Db    347 titssp-ssttlst------siptttpeitstlsslpdnaicsyldetttsttfttttml 399

Qy    485 ATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP 544
Db    400 tsttt-------------eepststtttevtstsstvttt-------------epttt 431

Qy    545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGG 604
Db    432 lttstaststtepststvttsps-------tspvtstvts------sssssttvttp--- 475

Qy    605 TDTVIIREPPNHTVTTTEYWSQSFATTTTVTGP------PSGTDTVIIREPPNPTVTTTE 658
Db    476 tsteststspsstvtts-----ttapststtgpsssssstpsstasssvsstasstqssts 530

Qy    659 YWSQSYATTT---TITAPPGETDTVLIREPPNHTVTTTEYWSQS-------------YAT 702
Db    531 --tqqsstttksetttssdgtnpdfyfve----katttfydstsvnltlnsglgiigyqt 584

Qy    703 TTTVTAPPGETDTVLIRE-----------------------PPNHTVTTT--------- 729
```



RESULT 7
AAY95556
AAY95556 standard; Protein; 3178 AA.
XX
AC AAY95556;
XX
DT 10-OCT-2000 (first entry)
XX
DE Caenorhabditis elegans LOV-1 (location of vulva) protein.
XX
KW Location of vulva; LOV-1 gene; nematode; mating behaviour;
KW polycystin; polycystic kidney disease; animal model;
KW signal transduction.
XX
OS Caenorhabditis elegans.
XX
PN WO200040711-A2.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00521.
XX
PR 06-JAN-1999; 99US-0115127.
XX
PA (CALY) CALIFORNIA INST OF TECHNOLOGY.

XX
PI Sternberg PW, Barr MM;
XX
DR WPI; 2000-452537/39.
DR N-PSDB; AAA50043.
XX
PT Isolated nucleic acid molecules from Caenorhabditis elegans useful for
PT producing transgenic nematodes with altered mating behavior for
PT identifying genes or regulatory factors involved in polycystic kidney
PT disease -
XX
PS Claim 5; Page 114-123; 142pp; English.
XX
CC The present sequence is that of the polypeptide encoded by the
CC Caenorhabditis elegans LOV-1 gene (see AAA50043), an orthologue of
CC the human polycystin-1 (PKD1) gene that is defective in human
CC autosomal dominant polycystic kidney disease (PKD). The LOV-1
CC protein is involved in chemosensory or mechanosensory signal
CC transduction in sensory neurons. It is required for 2 male
CC sensory behaviours, 'response' of males to hermaphrodites, and
CC 'location' of the vulva. Nematodes such as C. elegans that
CC express mutant or wild-type LOV-1 or PKD-2 (see AAY95557) can be used
CC to study the functions of the proteins encoded by these genes, to
CC screen for other genes involved in PKD, to identify mutations
CC involved in the disease, and to screen for drugs that affect PKD.
CC Behaviours controlled by the action of the genes or gene products
CC are identified and used in the assays. Hence, an animal model is
CC provided that permits study of the etiology of PKD and provides a
CC tool to identify the genes involved in the disease pathway, to
CC identify compounds that may be used to treat or alter the disease
CC progression, lessen its severity or ameliorate symptoms.
XX
SQ Sequence 3178 AA;

Query Match 8.5%; Score 553.5; DB 21; Length 3178;
Best Local Similarity 24.3%; Pred. No. 5.5e-20;
Matches 290; Conservative 154; Mismatches 401; Indels 349; Gaps 51;



```
Db    585 siectsptssnyvsttkdgacftksvsmprlggtypastfvgpgnytfratmttddkkvy 644

Qy    730 ----EYWSQSYATTT-----TVTAPPGGTDTVIIREPPNP-----TVTTTEYWSQSFATT 775
Db    645 ytyanvyiqeyssttiesesstsavasstsst----pstpsstlststvtepsstrssds 700

Qy    776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMN 835
Db    701 tttsa--gsttt--lqestttseesttdsstttisdtststsspssttadsts--tlsvd 754

Qy    836 TPTSISSDGMLLSST----------TLVTESETTTELICSDGKECSRLSS---------- 875
Db    755 qfdfildsglswnetrhnedsinivplptnaitpter--sqtfecrnvstepfliikest 812

Qy    876 ------------SSGIVTNP----------------------------------DSN 886
Db    813 clnysntvlnatyssnipiqpietflvgigtyefrinmtdlttmqvvshiftlnvvadst 872

Qy    887 ESSIVTSTVPT-ASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSS 945
Db    873 stsevtsttstgsssessaisttsgieststleasttdasqdsststsdsgtt-----sd 927

Qy    946 VTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTST------------EIEIVTT 992
Db    928 sttidssnstpstsdssg---lsqtpsdsssasdsmrtttvdpdastetpydfvlenltw 984

Qy    993 SSTK-------VLPPVVSS-----NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTG 1040
Db    985 netvyysenpfyitpipnkepgalttamtcqcrndssqpfvllkesnclte-------fg 1037

Qy   1041 DNGDNTSST--NPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGEN-SGLTTSTEIE 1097
Db   1038 kngaysasvsfnpmtsfvpat----------gtyeflinvtnrasgesashiftmnvvl 1086

Qy   1098 ATTTSPTEAPSPAVSS---------GTDVT--TEPTDTREQPTTLSTTSKTNSELVATTQ 1146
Db   1087 pttt--tetppttvsssddaggktggtgatggtggtgsggsattlst-----gdavrstt 1139

Qy   1147 ATNENGGKSPSTDLTSSLTT----GTSASTSANSELVTSGSVT----------------- 1185
Db   1140 sgsgsgqsstgsgaggsgttasgsgsggssgtgsdgvnsgkttalngdgtgsgtattpgs 1199

Qy   1186 ----GGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQV-----TSSSPSTNT 1230
Db   1200 hlgdggstsgsgsdsngssgvstkssgsdtsgssdssgangafsataqpstrt 1253
```

RESULT 8
AAY59288
ID AAY59288 standard; Protein; 957 AA.
XX
AC AAY59288;
XX
DT 25-APR-2000 (first entry)
XX
DE Human MUC11 polypeptide.
XX
KW Mucin; MUC11; MUC12; human; chromosome 7q22; epithelial inflammation;
KW Crohn's disease; ulcerative colitis; asthma; chronic bronchitis;
KW colorectal cancer; cystic fibrosis; inflammatory bowel disease;
KW breast cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 68
FT /note= "potential N-glycosylation site"
XX
PN WO200004142-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-AU00579.
XX
PR 16-JUL-1998; 98AU-0004708.
XX
PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
PA (ORDE-) ORDER OF SISTERS OF MERCY IN QUEENSLAND.
XX
PI Williams SJ, Antalis TM, Mcguckin MA, Gotley DC;
XX
DR WPI; 2000-182416/16.
DR N-PSDB; AAZ58824.
XX
PT Novel MUC nucleic acid corresponding to mucin gene, useful for treating
PT associated disease conditions e.g. colorectal, breast cancer, cystic
PT fibrosis and inflammatory bowel disease -
XX
PS Claim 10; Page 83-88; 103pp; English.
XX
CC The invention provides mucin genes (MUC11 and MUC12) located on human
CC chromosome 7q22. The mucin genes or its portion is used in detecting
CC polymorphism, mutation, deletion, truncation and expansion in the gene
CC or its gene transcript. Pharmaceutical compositions and gene therapy
CC constructs comprising the mucin genes are used for treating disease
CC conditions associated with aberrant Mucin expression, altered properties
CC of mucus or epithelial inflammatory processes involving Mucins like
CC Crohn's disease, ulcerative colitis, asthma, chronic bronchitis and
CC colorectal cancer, cystic fibrosis, inflammatory bowel disease and breast
CC cancer. The mucin genes and the polypeptides are used for determining
CC these diseases or their predisposition. The MUC11 and MUC12 polypeptides
CC are used for preparing antagonist and antibodies. The present sequence
CC represents the human MUC11 polypeptide.
XX
SQ Sequence 957 AA;

Query Match 8.4%; Score 545.5; DB 21; Length 957;
Best Local Similarity 25.9%; Pred. No. 3.2e-20;
Matches 266; Conservative 139; Mismatches 406; Indels 215; Gaps 43;

```
Qy    338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
Db      6 ttafpgsttmpgvsqestas--hsspgstdttlspgstta-----sslgpesttfhsgpg 58

Qy    398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
Db     59 stettllpdnttasglleastpvhsstgsphttl------spagsttrqgesttfqswpn 112

Qy    447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
Db    113 skdtt---pappttsafvelstttshgspsstptthfsasstt1grseesttvhsspvat 169

Qy    498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
Db    170 att-----psparstt---sglveesttyhsspgstqtmhfpesdttsgrgeesttshss 221

Qy    558 TTTTVTAPPGGTDSVIIREP------PNHTVTT----------TEYWSQSYA---TTTT 597
Db    222 tthtissap-sttsalveeptsyhsspgstatthfpdssttsgrseestashsnqdatgt 280

Qy    598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
Db    281 ivlparsttsvllges-----ttspissgsmettalpgstttpglsekstfhssprspa 335

Qy    654 VTTTEYWSQSYAT---TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
Db    336 ttlspasttssgvseesttshsrpgsthttafpds-----tttpglsrh---sttshssp 387

Qy    711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
Db    388 gstdttll--pas---tttsgpsqe---sttshsspgstdtal-----spgsttalsfgq 434

Qy    771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
Db    435 e---sttfhsspgsthttlfpdsttssgiveastrvhsstgsprttlspasstspglqge 491
```

```
Qy    823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
          ::   |  :|:  ||::       ||      |:||          :   :||   |
Db    492 stafqthpasthttpst--------pstatapveesttyhrspsstptthfpass---tt 540

Qy    883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
            :| | :  : | ||  | | : |:| |   :::  :|     :||     |   |||
Db    541 sghsekstifhsspdasgttpssahsttsgrgesttsrispgsteittlpgst---ttpg 597

Qy    941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
           || : |:       |: : | : :|    |: :| |  ||  ||   :   |:|    |
Db    598 -lseasttfyssprsptttlspasmtslgvge-esttsrsqpgsthstvspastt---tp 652

Qy   1001 VVS--SNTDLTSEPTNTR-------------EQPTTL-----STTSNSITEDITTS--- 1036
           :|  | |  :| | :|              |:|||      || :   ||| |||
Db    653 glseesttvyssspgstettvfprstttsvrgeepttfhsrpasthttlftedsttsglt 712

Qy   1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
           : |   |   |:   :|  || | |   ||      |  |  |:| |:    |||
Db    713 eestafpgspastqtglp--atlttadlgeesttfpsssgstgttlsparsttsglvges 770

Qy   1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
                            || :  | :||      ||    ||| :: : |: | : :
Db    771 tpsrlspsstetttlpgspttpslseksttfytsprspdatlspatttssgvseesstsh 830

Qy   1126 EQPTTLSTTS-------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
           || :  ||:              ||:    :|:||     || :  |||
Db    831 sqpgsthttafpdstttsglsqepktshssqgsteat-----lspgsttasslgqqsttf 885

Qy   1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
            : | : :|    : :||| |       :|    |:|     |::      ::|  |
Db    886 hsspgdtettllpddtitsglveastpthsstgslhttltpasstsaglqeesttfqswp 945

Qy   1223 SSSPST 1228
          ||| :|
Db    946 sssdtt 951
```

RESULT 9
AAM24513
ID AAM24513 standard; Protein; 957 AA.
XX
AC AAM24513;
XX
DT 12-OCT-2001 (first entry)
XX
DE C900P predicted amino acid sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 437-440; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 957 AA;

Query Match 8.4%; Score 545.5; DB 22; Length 957;
Best Local Similarity 25.9%; Pred. No. 3.2e-20;
Matches 266; Conservative 139; Mismatches 406; Indels 215; Gaps 43;

```
Qy    338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
          |||        | | : |       | :| ||   | ||:     : :|  :|      |
Db      6 ttafpgsttmpgvsqestas--hsspgstdttlspgstta-----sslgpesttfhspgpg 58

Qy    398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
            | ||:  : |   | :  :|   : | |  | :      :|  |||         : |
Db     59 stettllpdnttasglleastpvhsstgsphttl------spagsttrqgesttfqswpn 112

Qy    447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
          |  ||    |||  |   :     :|   : | | ::| |  |      :|||  : |  |
Db    113 skdtt---pappttsafvelsttshgspsstptthfsassttlgrseesttvhsspvat 169

Qy    498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
            :      |:|  :||   |     :||  : || | ::   |     :     |  :   :
Db    170 att-----psparstt---sglveesttyhsspgstqtmhfpesdttsgrgeesttshss 221

Qy    558 TTTTVTAPPGGTDSVIIREP------PNHTVTT----------TEYWSQSYA---TTTT 597
          || |::: |  | | :: ||      |  | ||           :|  : |::    | |
Db    222 tthtissap-sttsalveeptsyhsspgstatthfpdsttsgrseestashsnqdatgt 280

Qy    598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
          :  |   | :|:: |      ||:   | |  ||    :| |   |   |     | :|
Db    281 ivlparsttsvllges-----ttspissgsmettalpgstttpglsekstfhssprspa 335

Qy    654 VTTTEYWSQSYAT---TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
          | :   : |      :||  : || | |   :      |||   |:    :||  : |
Db    336 ttlspasttssgvseesttshsrpgsthttafpds-----tttpglsrh---sttshssp 387

Qy    711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
          | ||| |:  | :   |||   ||    :||  : || ||| :     :|  ||   : |
Db    388 gstdttll--pas---tttsgpsqe---sttshsspgstdtal-----spgsttalsfgq 434

Qy    771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
            :||  : || | | :  :| :|| |        |:: :  |:: |: |   :
Db    435 e---sttfhsspgsthttlfpdsttssgiveastrvhsstgsprttlspasstspglqge 491
```

```
Qy     823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
           ::   |  :|:  ||::         ||      |:||       :   :||   |
Db     492 stafqthpasthttpst--------pstatapveesttyhrspsstpthfpass---tt 540

Qy     883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
             :| | :  : | ||  | | : |:| |   :::  :|     :||    |   |||
Db     541 sghsekstifhsspdasgttpssahsttsgrgesttsrispgsteittlpgst---ttpg 597

Qy     941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
            || : |:       |: : | : :|    |: :| |  ||  ||   :   |:|    |
Db     598 -lseasttfyssprspptttlspasmtslgvge-esttsrsqpgsthstvspastt---tp 652

Qy    1001 VVS--SNTDLTSEPTNTR-------------EQPTTL-----STTSNSITEDITTS--- 1036
           :|  | |  :| | :|               |:|||      || :   ||| |||
Db     653 glseesttvysspgstettvfprstttsvrgeepttfhsrpasthttlftedsttsglt 712

Qy    1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
            : |   |   |:   :|  || | |   ||      |  |  |:| |:    |||
Db     713 eestafpgspastqtglp--atlttadlgeesttfpsssgstgttlsparsttsglvges 770

Qy    1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
                           || :  | :||      ||    ||| :: : |: | : :
Db     771 tpsrlspsstetttlpgspttpslsekstfytsprspdatlspatttssgvseesstsh 830

Qy    1126 EQPTTLSTTS-------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
           || :  ||:              ||:    :|:||     || :   |||
Db     831 sqpgsthttafpdstttsglsqepktshssqgsteat-----lspgsttasslgqqsttf 885

Qy    1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
             : | : :|    : :||| |       :|    |:|    |::      ::|  |
Db     886 hsspgdtettllpddtitsglveastpthsstgslhttltpasstsaglqeesttfqswp 945

Qy    1223 SSSPST 1228
           ||| :|
Db     946 sssdtt 951
```

RESULT 10
AAW06725
ID AAW06725 standard; Protein; 849 AA.
XX
AC AAW06725;
XX
DT 07-FEB-1997 (first entry)
XX
DE FLO1 protein, involved in flocculation, derived from S.cerevisiae.
XX
KW Flocculation; aggregation; floating; yeast; beer; Saccharomyces;
KW fermentation; ds.
XX
OS Saccharomyces cerevisiae.
XX
FH Key Location/Qualifiers
FT CDS 1..2550
FT /*tag= a
FT /note= "partial sequence only, no start codon"
XX
PN JP08205900-A.
XX
PD 13-AUG-1996.
XX
PF 01-FEB-1995; 95JP-0015453.
XX
PR 01-FEB-1995; 95JP-0015453.
XX
PA (KIRI) KIRIN BREWERY KK.
XX
DR WPI; 1996-419830/42.
DR N-PSDB; AAT45666.
XX
PT DNA molecule involved in yeast flocculation - useful for determining
PT if a sample yeast has a flocculation activity
XX
PS Claim 1; Page 6-10; 16pp; Japanese.
XX
CC AAW06725 is the FLO1 protein derived from a partial gene sequence
CC from Saccharomyces cerevisiae. The gene was named Lg-FLO1. The
CC FLO1 protein is believed to be involved in yeast flocculation
CC (aggregation). The DNA and protein are used in a method for
CC determining whether or not a sample yeast has flocculating activity,
CC i.e. to distinguish if a yeast is a flocculating or floating yeast.
XX
SQ Sequence 849 AA;

Query Match 8.2%; Score 534; DB 17; Length 849;
Best Local Similarity 26.0%; Pred. No. 1.1e-19;
Matches 244; Conservative 152; Mismatches 386; Indels 156; Gaps 39;

```
Qy     380 TKTAPIGETATVIVDVP--YHTTTTVTSEWTGTIT------TTTTRTN--PTDSIDTVVV 429
           |   |  ||  ::|  |   :|  | |  ||||  |      || | ||  |||    :|:
Db       2 tngqptdet-vivvktpttantivtttkpwtgtftststemttvtgtngqptde-tvivi 59

Qy     430 QVPLPNPTVS-TTEYWSQSFATT----TTVTAPPG-GTD--TVIIREPPNH--TVTTTEY 479
           : |     :| ||| |: :| :|    ||:|   |  ||   ::|| | :     ||||
Db      60 rtptseglistttepwtgtftststevttttgtngqptdetvivirtptseglvtttep 119

Qy     480 WSQSFATT----TTVTAPPG--GTDSVIIREPPNPTVTTTEYWSQSFATTTTVTA----- 528
           |: :| :|    ||:|   |    ::||: : |   ::::    | |   |:::|:
Db     120 wtgtftststgmttitgtnglptdetvivvktpttaisssslslsssggqitssitssrpil 179

Qy     529 ----PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTT 584
               |  ||   :|      :  |:   : |  |:: ||:      |||      :  ||:
Db     180 tpfypsngts--visssdtsssdtsslvtsslvtsslvts------svisssvtsslvts 231

Qy     585 TEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTV 644
           :     |  |:: ||:       :||      :   |::   | |  :::|       | |
Db     232 svi--sssvtsslvts------svisssvtsslftsspvisssvissstt------tsts 277

Qy     645 IIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT 704
           |  |    :|  |   :   : : | :|    : : :  | |  | :::     : |||:
Db     278 iffesskssviptssstsgsseskt ssassssssssissespkstyssslppvtsatts 337

Qy     705 TVTAP--PGETDTVLIRE-----PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREP 757
             ||   |  |     :|     ||  : ||::  : |    |: |       ::    |
Db     338 qetasslppvtsvttsqeitsslppvistttsqetasslppvtsattsqetassl----p 393

Qy     758 PNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPH 817
           |  :|||::  : |    |:||       :  |  |:     :|:| :||| :|
Db     394 pvtsvttsqeitsslppvtsvtt------sqeitsslppvtsattsqeitsslpp----- 442

Qy     818 YVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSS 877
           | | |:   |  :|||:  || ::     ||   ||   |: |           ::||
Db     443 -vtsattsqet--asslppvtsaatsqetasslppvtsvttsqe----------itssl 488

Qy     878 GIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGI-----SATSSDNVSKSGVSVTTETSV 932
             ||:  :::   :||::|  :::: |   |   :     : || :  |    : ||:||
Db     489 ppvtsvttsqe--itsslppvtsvttsqeitsslppvtsattsqetasslppatttktse 546

Qy     933 TTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTT 992
           |   |     | | : :  |:| | :      | | :|     |     ::|||   ||
Db     547 qttlvtvtsceshvctesissaivsta------tvtvsgvtteyttwcpistteitkqtt 600

Qy     993 SSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPV 1052
           :||          | : | | :| ||: | |:  ::  : :          :|:: | |
Db     601 ettkqtkgttkqtkgtteqttettkq-ttvvtisscesdicsktaspaivststatindv 659

Qy    1053 PTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP---SP 1109
           |  | |    |  ::|               : : | | |  |:   | | :|    |
Db     660 tteyt-twcpisttesk---------------qqttlvtvtscesgvcsetaspaivst 702

Qy    1110 AVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSL 1164
           | ::  || |      |  | ||    | :|| ||   ||:: |   |     | :||||
Db     703 atatvndvvtvyptwrpqttneq----svsskmns---atsetttntgaaetktavtssl 755
```

```
Qy   1165 TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSS 1224
Db    756 srfnhaetqtasatdvigh--sssvvsvsetgntkglitselstmsqqprstpastmvgs 813

Qy   1225 SPSTNTFIASTYDGSGSIIQHSTWLYGLITL--LSLFI 1260
Db    814 --stasleistyvgian--------glltnngisvfi 840
```

RESULT 11
AAR47578
ID AAR47578 standard; Protein; 894 AA.
XX
AC AAR47578;
XX
DT 19-JUL-1994 (first entry)
XX
DE Flocculation protein of Saccharomyces cerevisiae.
XX
KW Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW inulinase; alpha-amylase; Saccharomyces cerevisiae;
KW floculation protein; enzymatic process; fermentation;
KW biodegradation; catalysis.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9401567-A.
XX
PD 20-JAN-1994.
XX
PF 07-JUL-1993; 93WO-EP01763.
XX
PR 08-JUL-1992; 92EP-0202080.
PR 14-DEC-1992; 92EP-0203899.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
PI Klis FM, Schreuder MP, Toschka H, Verrips CT;
XX
DR WPI; 1994-035071/04.
DR N-PSDB; AAQ54029.
XX
PT Immobilisation of enzymes to microbial cell wall - by prodn. of
PT fusion protein of enzyme linked to anchoring protein
XX
PS Example 10; Page 59-64; 99pp; English.
XX
CC The flocculation protein is used in a method to immobilise enzymes
CC to a microbial cell wall. The coding sequence (FLO 1) is used in
CC the production of a recombinant polynucleotide which comprises a
CC structural gene encoding a protein with catalytic activity and at
CC least part of a gene encoding at least the C-terminus of a protein
CC capable of anchoring in a eukaryotic or prokaryotic cell wall. The
CC anchoring fragment or protein is selected from alpha agglutinin,
CC AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a
CC proteinase of lactic acid bacteria. The recombinant polynucleotide
CC preferably also comprises a sequence encoding a signal peptide to
CC ensure secretion of the expressed product. The signal peptide is
CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC alpha-amylase of Bacillus or proteinases of lactic acid bacteria.
CC The host microorganism can be used for performing enzymatic
CC processes on an industrial scale.
XX
SQ Sequence 894 AA;

Query Match 7.9%; Score 516; DB 15; Length 894;
Best Local Similarity 26.3%; Pred. No. 9.8e-19;
Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

```
Qy    437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
Db     53 tysnaaymaygyasktklgsvggqtdisidynipcvsssgtfpcpqedsygnwgckgmga 112

Qy    478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
Db    113 csnsqgiaywstdlfgfytt--------------ptnvtlemtgyflppqtgsytfkfa 157

Qy    522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
Db    158 tvddsailsvggatafnccaqqqppitstnf-------tidgikpwggs------lppni 204

Qy    581 TVTTTEYWSQSY-------------ATTTTVTAPPGGT------------------DTVI 609
Db    205 egtvymyagyyypmkvvysnavswgtlpisvtlpdgttvsddfegyvysfdddlsqsnct 264

Qy    610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
Db    265 vpdpsnyavsttttttepwtgtftststemttvtgtngvptdetvivirtptseglistt 324

Qy    657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
Db    325 tepwtgtftststevttitgtngqptdetvivirtptseglistttepwtgtftststem 384

Qy    704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
Db    385 ttvtgtng--------qptdetv-------------------------ivirtptseglv 411

Qy    762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
Db    412 ttttepwtgtftststemstvtgtnglptdetvivvktpttaissslsssssgqitssit 471

Qy    812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
Db    472 s-srpiitpfypsngtsvisssvisssvtsslftsspvi--sssvissstttstsifses 528

Qy    868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
Db    529 skssviptssstsgssesetss-----------agsvssssfissesskptyssslp 576

Qy    928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
Db    577 lvtsattsqetasslppatttkt--------seqttlvtv--------------tsces 613

Qy    988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
Db    614 hvctes---ispaivstatvtvsgvtteyttwcpistte-------ttkqtkgtt-eqtt 662

Qy   1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
Db    663 ettkqttvvt---isscesdvcsktaspaivststatingvtteyttwcpisttesrqqt 719

Qy   1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
Db    720 tlvtvtscesgvcsetaspaivstatatvndvvtvyptwrpqtaneesvsskmnsatget 779

Qy   1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
Db    780 t--tntlaaetttntvaaetitntga---aetktvvtsslsrsnhaetqtasatdvighs 834

Qy   1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
Db    835 ssvvsvsetgntksltssglstmsqqprstpassmvgystasleistyagsat 887
```

RESULT 12
AAR58754
ID AAR58754 standard; Protein; 894 AA.
XX
AC AAR58754;
XX
DT 27-MAR-1995 (first entry)

XX
DE S. cerevisiae FLO1.
XX
KW Binding protein; immobilization; chimeric protein;
KW anchoring protein; Saccharomyces cerevisiae; flocculation;
KW FLO1 gene; scFv; single chain antibody; monoclonal antibody;
KW MAb; human chorionic gonadotropin; HCG.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9418330-A.
XX
PD 18-AUG-1994.
XX
PF 10-FEB-1994; 94WO-EP00427.
XX
PR 10-FEB-1993; 93EP-0200350.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
PI De Geus P, Frenken LGJ, Klis FM, Toschka HY, Verrips CT;
XX
DR WPI; 1994-279751/34.
DR N-PSDB; AAQ67360.
XX
PT Immobilised binding proteins for specific cpds - obtd. by
PT expressing chimeric proteins comprising the binding protein and a
PT cell wall-anchoring protein in host cells
XX
PS Disclosure; Page 36-39; 78pp; English.
XX
CC Valuable compounds are isolated from complex mixtures by use of
CC immobilized ligands composed of an anchoring protein and a binding
CC protein. A suitable anchoring protein is yeast FLO1 (associated
CC with flocculation), and a gene encoding a chimeric scFv-FLO1 protein
CC that will anchor in the cell wall of a lower eukaryote, and which
CC binds HCG with high specificity, was produced.
XX
SQ Sequence 894 AA;

Query Match 7.9%; Score 516; DB 15; Length 894;
Best Local Similarity 26.3%; Pred. No. 9.8e-19;
Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

```
Qy    437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
          | |   | :  :|: | : :  | ||  |    |  : : |
Db     53 tysnaaymayggyasktklgsvggqtdisidynipcvsssgtfpcpqedsygnwgckgmga 112

Qy    478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
                  |||   |   ||               | | |:  | |        ::  ||
Db    113 csnsqgiaywstdlfgfytt--------------ptnvtlemtgyflppqtgsytfkfa 157

Qy    522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
          |   :     ||  :        | :|:| :       |     | ||:       |||
Db    158 tvddsailsvggatafnccaqqqppitstnf-------tidgikpwggs------lppni 204

Qy    581 TVTTTEYWSQSY------------ATTTTVTAPPGGT-----------------DTVI 609
            |   |   |                :|| | | |
Db    205 egtvymyagyyypmkvvysnavswgtlpisvtlpdgttvsddfegyvysfdddlsqsnct 264

Qy    610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
          : :| |:    | |||| |: :| :|    |||||    |:    ::|| | :  :  ||
Db    265 vpdpsnyavsttttttepwtgtftststemttvtgtngvptdetvivirtptseglistt 324

Qy    657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
          || |: :: :|    ||||   |: ||   ::|| | :  :  |||| |: :: :|
Db    325 tepwtgtftststevttitgtngqptdetvivirtptseglistttepwtgtftststem 384

Qy    704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
          ||||   |        :| : ||                          ::|| | :
Db    385 ttvtgtng--------qptdetv------------------------ivirtptseglv 411

Qy    762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
          ||||  |: :| :|:|  :   ||      :|||: :    ::|||  |:||  ||| |
Db    412 ttttepwtgtftststemstvtgtnglptdetvivvktpttaissslsssssgqitssit 471

Qy    812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
          | ||    | | :: || :|:   ||  | :  :|  ::  |:::::| ||:  | |:
Db    472 s-srpiitpfypsngtsvisssvisssvtsslftsspvi--sssvissstttstsifses 528

Qy    868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
          : | : :||    : :|  ||            : |:||:  ||: || : : |  |:
Db    529 skssviptssstsgssesetss-----------agsvssssfissessksptyssslp 576

Qy    928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
            || || | | : |  : |: |        ||  : ||                || |
Db    577 lvtsattsqetasslppatttkt--------seqttlvtv--------------tsces 613

Qy    988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
          : | |   : | :||: |   |  |        :|||        || |  |   : |:
Db    614 hvctes---ispaivstatvtvsgvtteyttwcpistte------ttkqtkgtt-eqtt 662

Qy   1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
            |   || |   |: | |  :|: |  |||    |  :  ||    :||    |
Db    663 ettkqttvvt---isscesdvcsktaspaivststatingvtteyttwcpisttesrqqt 719

Qy   1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
          |  |          |  ||| ||: |      ||  ||   :     | :|| ||    |
Db    720 tlvtvtscesgvcsetaspaivstatatvndvvtvyptwrpqtaneesvsskmnsatget 779

Qy   1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
          |  ||    :: :  : :   | | |  | :: | : |::         |||::   ||
Db    780 t--tntlaaetttntvaaetitntga---aetktvvtsslsrsnhaetqtasatdvighs 834

Qy   1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
          :||     | :   ::::  :|:|||  |:          || :   ||| || :
Db    835 ssvvsvsetgntksltssglstmsqqprstpassmvgystasleistyagsat 887
```

RESULT 13
AAB40945
ID AAB40945 standard; Protein; 1532 AA.
XX
AC AAB40945;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF709 polypeptide sequence SEQ ID NO:1418.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC75154.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 1203-1206; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1532 AA;

Query Match 7.8%; Score 507; DB 21; Length 1532;
Best Local Similarity 23.3%; Pred. No. 5.6e-18;
Matches 324; Conservative 200; Mismatches 529; Indels 338; Gaps 62;

```
Qy      55 SLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFST--LTC 112
           |   |  |||:| :: :         ||   | :  :|   ||    | :: | ||    |
Db     274 stgstlgnpgetssvpvtgslmpvtsaalvtvdpegqspatfsrtstqdttafsknhqtq 333

Qy     113 TVNDALKSSIKAFGTVT---------LPIAFNVGGTGSSTDLEDSKCFTAGTNTV----- 158
           :|     | |    |:|          |  ||  || |       : | :|  |:
Db     334 svettrvsqintlntltpvttstvlsspsgfnpsgtvs------qetfpsgettisspss 387

Qy     159 ---TF-----------------NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLN--- 194
               ||                 |  :  :|:       |   |   ::  :  :|:
Db     388 vsntflvtskvfrmpisrdstlgnteetslsvsgtisaitskvstiwwsdtlstalspss 447

Qy     195 ---KVTTLFVAPQCENGYTSGTMGFSSS-NGDVAIDCSNIHIGIT-------KGLNDWNY 243
               |::| |   | |   |:|     || :  |: :   :|   |       ||   |:
Db     448 lppkistafhtqqsegaettgrpherssfspgvsqeiftlhetttwpssfsskghttwsq 507

Qy     244 P--VSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTC--- 298
           |: : : |:  | |          ||  | : : :||        |::  |
Db     508 telpststgaatrlvtgnp-----stgaagtiprvpskvsaigepgepttysshsttlpk 562

Qy     299 ---AGSRLQSKPFTLRWTGYKNSDAGSNGIVIVAT-TRTVTDSTTAVTTLPFNPSVDKTK 354
               ||:: |       ||     : |: |  :::: : :||      |:   :| :|:
Db     563 ttgagaqtq-------wt----qetgttgeallsspsysvtqmiktatspssspmldrht 611

Qy     355 TIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHT----------TTTVT 404
           : :|     ||  :|::   :| : | |:|    |   |    ||          ||
Db     612 sqqi-----ttapstnh---stihststspqespa---vsqrghtqapqttqesqttrsv 660
```

```
Qy     405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV 464
           | | | | ||  :: | |  :    ||   ||:|         | ||   || |
Db     661 spmtdtktvttpgssftasghspseivpqdaptis---------aattfapaptg----- 706

Qy     465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT 524
                 ||           | || : | |   |: :      |:  |:   : :
Db     707 -----dghttq---------apttalqatpsshdatl-----gpsggtslsktgaltlan 747

Qy     525 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTT 584
           :| : ||| :                  |: : |:|:  ||                :|
Db     748 svvstpggpeg--------------qwtsasastspdtaa--------------amth 777

Qy     585 TEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTV 644
           |     : |:  | |: |  : :          |   |   | | |: ||    |||:: :
Db     778 thqaesteasgqtqtsepassgs-------rttsagtatpsssgasgtt----psgsegi 826

Qy     645 -----IIREPPNP------TVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTT 693
                |   ||       | :|||  | | |:   |    |   :::   |     | :
Db     827 stsgettrfssnpsrdshttqsttellsas-ashgaipvstgmassiv---pgtfhptls 882

Qy     694 EYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT-----TVTAPPGG 748
           |         :|   | |::          |  |   :  :|:  | |     |::
Db     883 e--------astagrptgqssptspsaspqetaaisr-maqtgrtrtsrgsdtislasqa 933

Qy     749 TDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGG------------TDTVIIYESMS 795
           |||     |  |::|::   |     | | |  | |              |   : | | :
Db     934 tdtfstvpptppsitssgltspq---tqthtlspsgsgktfttalisnatplpvtyassa 990

Qy     796 SSKISTSSN--DITSIIPSFSRPHYVNSTTSDLSTFESSSM--NTPTSISSDGMLLSSTT 851
           |:  :|  :  | :|:    : |  | | :| :|| ::: :   :|:| ||      :|:
Db     991 stghttplhvtdassvstghatplpvtspss-vstgdttplpvtspssassg----hats 1045

Qy     852 L-VTESET-----TTELICSDGKECS-----------RLSSSSGIVTN-PDSNESSIV-- 891
           | ||:: :      | |  :|   |                |:|:|  |: | :: ||:
Db    1046 lpvtdasslstghatslhvtdassvstghatllhvtdassastghttslpvtdassvstg 1105

Qy     892 -TSTVPTASTMSDSLSSTDGISATSSDNVS---KSGVSVTTETSVTTIQTTPNPLSS--- 944
            |: :|   | | |   |  :  | : :||  ' : : ||: :|||:|  ||| |::|
Db    1106 dttplpvtdtssastgdttplhvtdassvstghatplhvtslssvstgdttplpvtspss 1165

Qy     945 ----SVTSL--TQLSSIPSVSESESKVTFTSNGDNQSGT-------------HDSQSTST 985
                |||  |  ||: :   :   ||  |:  :  |                | :   |
Db    1166 assghatslpvtdassvstghatslpvtipssassgdatslpvtslsslstghatplpvt 1225

Qy     986 EIEIVTTSSTKVLPPVVSSN-----------TDLTSEPTNTREQPTTLSTTSNSITEDIT 1034
           :   :|     ||   :|:           || :|  |      |  ::  |:  | | |
Db    1226 slssastghatplpvtdtssvstghatslltvdassvstg-hatplhvtdassvstgdtt 1284

Qy    1035 TSQPTGDNGDNTSSTNPVPTVATSTLAS-------ASEEDNKSGSHESASTSLKPSMGEN 1087
              |  :  :|  | |:|   ||::::        ::  : | ||   :|||   : :
Db    1285 plpvtspssastgdttplpvtdtssvstgdttpllvtdtssvstsh---atsl--pvtdt 1339

Qy    1088 SGLTTSTEIEATTTSPTEA------PSP----AVSSGTDVTTEPTDTREQPTTLSTTSK 1136
           | ::||       | |: |      | |     :||:|   :   |||    :: ||
Db    1340 ssvstshatslpvtdpssastgdttplpvtdtssvstghatslpvtdt----ssastgdt 1395

Qy    1137 TNSELVATTQATNENGGKSPSTDLTSSLTTG----------TSASTSANSEL-VTSGS-- 1183
           |:  :  |: |:  :    | || ||| :||'          :||||   : | ||| |
Db    1396 tslpvtdtssastghatplpvtd-tssastghatpllvtdtssastghttplhvtspssa 1454

Qy    1184 VTGGAV-------ASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTY 1236
            || |        :|||   : |  ||:::|  :      |    :|:|    | :  |
Db    1455 stghatplpvtspssastshatslpvtdtssastghatpllvtdtssastghatpllvt- 1513

Qy    1237 DGSGSIIQHST 1247
           | | :   |:|
Db    1514 dtssastghat 1524
```

RESULT 14
AAW48299
ID AAW48299 standard; Protein; 1721 AA.
XX

```
AC   AAW48299;
XX
DT   17-AUG-1998  (first entry)
XX
DE   Cryptosporidium parvum GP900 antigen.
XX
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; 3' region.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   Domain          175..423
FT                   /note= "NINC domain"
XX
PN   WO9806430-A1.
XX
PD   19-FEB-1998.
XX
PF   11-AUG-1997;   97WO-US14104.
XX
PR   14-AUG-1996;   96US-0700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Gut J,  Leech J,  Nelson RC,  Petersen C;
XX
DR   WPI; 1998-159290/14.
DR   N-PSDB; AAV20700, AAV20701.
XX
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
XX
PS   Claim 27; Pages 65-68; 89pp; English.
XX
CC   The sequence is that of the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
XX
SQ   Sequence   1721 AA;
```

Query Match 7.5%; Score 488; DB 19; Length 1721;
Best Local Similarity 22.5%; Pred. No. 6.2e-17;
Matches 303; Conservative 128; Mismatches 469; Indels 448; Gaps 58;

```
Qy    214 MGFSSSNGDVAIDCSNIHIGITKG-LNDWNYPVSSESFSYTKTC---------TSNGIQI 263
Db     19 lnftsttgfttdtsmnwpvsitsgelkd---pnkqatisgsrscgwkqgysidsstgfrv 75

Qy    264 ---------KYQNVPAGYRPFIDAYIS----ATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
Db     76 dsitglptdpysncp--fnpvtgnlvsrstgktipntyagvyrsnet-----kttepsan 128

Qy    311 RWTG-YKNSD-------AGSNGIVI---------------------------------- 327
Db    129 tyagvyrsnetkttepsantnfllvdpkinapcnsensfeqgqifdmgskvyipytkcvg 188

Qy    328 ----------VATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTS 377
Db    189 vkhttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 248

Qy    378 YLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPT 437
Db    249 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttkkpt 308

Qy    438 VSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGT 497
```

```
Db    309 ttttt-ttttttttttttttttttttttttkkptttttttt--------ttttttttttt 359

Qy    498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
Db    360 ----------ttttttttttttttttttkptttttttttttkkptttt-------a 401

Qy    558 TTTTVTAPPGGTDSVIIRE-----------------------------------PPNH 580
Db    402 ttttttse---tesvikpdewcwlekngeceakgatyvgvigkdgriengmaftmipndd 458

Qy    581 TVTTTEYWSQSYATTTTV-----------------TAPP------------------- 602
Db    459 thvrfrfkvkdvgntisvrcrkgagklefpdrsldftippvaghnscsiivgvsgdgkih 518

Qy    603 ----GGTDTVIIREP--PNH-------TVTTTEY--------WSQSFAT------TTTVT 635
Db    519 vspygskdvslisapiqpselfnevycdtctakygaihsgyqtsadfvttttakpttttt 578

Qy    636 GPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEY 695
Db    579 gapggptttttgspskpttttt-----tkatttttlnp-------------iittt-- 617

Qy    696 WSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIR 755
Db    618 --tqkpttttttkvpg--------kppiattttt---lkpivttttttkatttttttv--- 661

Qy    756 EPPNPTVTTTEYWSQSFATTTTVTAPPGG-------------TDTVIIYESMSSSKIST 801
Db    662 ----pttttt--krdemtttttplpdigdieitpipiekmldkytrmiydynsgllds 715

Qy    802 SSN-----------DITSIIP----------------------SFSRPH-YVNSTTSDLS 827
Db    716 ndepipgsqagqiadtsnlfpvqthkstglpidpmvglpfdpksgnlvhpytnqtmsgls 775

Qy    828 TFESSSMNTPTSISSD-----------GMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
Db    776 vsylaakn--ltvdtdetyglpidtltgypldpvslipfnpetgelfdpisdeimn-gti 832

Qy    877 SGIVTNPDSNES--SIVTSTVPTASTM-----SDSLSSTDGI----------SATSSDNV 919
Db    833 agivsgisasesllsqksalidpatnmvvgefggllnpatgvmipgflgpseqtqfspei 892

Qy    920 SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSES----ESKVTFTSN--GDN 973
Db    893 edggi-------------ippevaaanadkfklsippsvpesipekdqkidsiselmydi 939

Qy    974 QSGTHDSQ-----------------STSTEIEIVT------TSSTKVLPPVVSSNTDLT 1009
Db    940 esgrligqvskrpipgsiagdlnpimktptqtdsvtgkpidpttglpfnpp-----tghl 994

Qy   1010 SEPTNTREQPTTLS-----TTSNSITEDITTSQPTGD-NGDNTSSTNPVPTVATS----T 1059
Db    995 inptnnntmdssfagaykyavsngiktdnvyglpvgeitglpkdpgsdipfnsttgelvd 1054

Qy   1060 LASASEEDNKSGSHESASTSLKPSMGENSGL---TTSTEIEATTTSPTEAPSPAVSSGTD 1116
Db   1055 pstgkpinnstagivsgkpglppiedengnlfdpstnlpidgnnqlvnpetnstvsgsts 1114

Qy   1117 VTTEP--------------TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTS 1162
Db   1115 gttkpkpgipvngggvvpdeeakdq-----adkgkdglivpptnsin----kdpvtntqy 1165

Qy   1163 SLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
Db   1166 snttg----niinpe---tgkvipgslpgslnypsf-----------ntpqqt--deit 1204

Qy   1223 SSSPSTNTFIASTYDGS-GSIIQHSTWL 1249
Db   1205 gkpvdtvtgl--pydpstgeiidpatkl 1230
```

RESULT 15
AAB11727

```
ID   AAB11727 standard; Protein; 1721 AA.
XX
AC   AAB11727;
XX
DT   28-OCT-2000  (first entry)
XX
DE   Portion of Cryptosporidium parvum NINC isolate GP900.
XX
KW   GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide.
XX
OS   Cryptosporidium parvum.
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-0928361.
XX
PR   13-SEP-1996;   96US-0026062.
PR   01-JUN-1993;   93US-0071880.
PR   29-MAY-1992;   92US-0891301.
PR   03-APR-1995;   95US-0415751.
PR   14-AUG-1996;   96US-0700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   N-PSDB; AAA61848, AAA61849.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and
PT   diagnosing or treating Cryptosporidium infections by competitive
PT   inhibition of the function of GP900 -
XX
PS   Claim 2; Column 61-70; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
     antibodies also inhibit sporozoite or merozoite attachment/invasion and
     additionally inhibit the binding of GP900 ligands to GP900. GP900
     proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents a portion of the GP900 protein of the NINC isolate of
CC   Cryptosporidium parvum.
XX
SQ   Sequence   1721 AA;
```

```
  Query Match             7.5%;  Score 488;  DB 21;  Length 1721;
  Best Local Similarity   22.5%;  Pred. No. 6.2e-17;
  Matches  303;  Conservative  128;  Mismatches  469;  Indels  448;  Gaps   58;

Qy     214 MGPSSSNGDVAIDCSNIHIGITKG-LNDWNYPVSSESFSYTKTC---------TSNGIQI 263
           :  |:|: |       |  : || | | |    |    : | :::|         :| | ::
Db      19 lnftsttgfttdtsmnwpvsitsgelkd---pnkqatisgsrscgwkqgysidsstgfrv 75

Qy     264 ---------KYQNVPAGYRPFIDAYIS----ATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
                   | | |  : |     :|      |  | |   | :: |        ::|
Db      76 dsitglptdpysncp--fnpvtgnlvsrstgktlpntyagvyrsnet-----kttepsan 128

Qy     311 RWTG-YKNSD-------AGSNGIVI---------------------------------- 327
            : | |::::       | :| :::
Db     129 tyagvyrsnetkttepsantnfllvdpkinapcnsensfeqgqifdmgskvyipytkcvg 188

Qy     328 ----------VATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTS 377
                     || | | :||  ||      :   | |         ||| ||:    ||:
Db     189 vkhttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 248

Qy     378 YLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPT 437
             | |      | |         |||| |:  | | ||||| |  | :  |          ||
Db     249 ttttttttttttttttttttttttttttttttttttttttttttttttttttttkkpt 308

Qy     438 VSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGT 497
            :||   :  |||| |      | |   ::|    | |||         |||| |      |
Db     309 ttttt-ttttttttttttttttttttttkkpttttttt--------ttttttttttt 359

Qy     498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
                     | |||   : :  |||| | |   | :         || |||       |
Db     360 ----------ttttttttttttttttttttkptttttttttttttkkpttttt-------a 401

Qy     558 TTTTVTAPPGGTDSVIIRE-----------------------------------PPNH 580
           |||| |:    |:|||  :                                     | :
Db     402 ttttttse---tesvikpdewcwlekngeceakgatyvgvigkdgriengmaftmipndd 458

Qy     581 TVTTTEYWSQSYATTTTV-----------------TAPP-------------------- 602
           |    :  :    | :|                  | ||
Db     459 thvrfrfkvkdvgntisvrcrkgagklefpdrsldftippvaghnscsiivgvsgdgkih 518

Qy     603 ----GGTDTVIIREP--PNH-------TVTTTEY--------WSQSFAT------TTTVT 635
                |  |  :|  |  |:           | :|         |  | |      ||| |
Db     519 vspygskdvslisapiqpselfnevycdtctakygaihsgyqtsadfvttttakpttttt 578

Qy     636 GPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEY 695
           | |    |      |  || |||     : ||||| |  |                :|||
Db     579 gapgqpttttttgspskpttttt-----tkatttttttlnp-------------iittt-- 617

Qy     696 WSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIR 755
                |||| |  ||          :||  | |||     :   ||||  |    | ||
Db     618 --tqkpttttttkvpg--------kppiattttt---lkpivtttttkatttttttv--- 661

Qy     756 EPPNPTVTTTEYWSQSFATTTTVTAPPGG--------------TDTVIIYESMSSSKIST 801
               || |||    :   ||||   | |                | :||:  |   : :
Db     662 ----pttttttt--krdemtttttplpdigdieitpipiekmldkytrmiydynsgllds 715

Qy     802 SSN-----------DITSIIP----------------------SFSRPH-YVNSTTSDLS 827
           :             | ::: |                    | :  | | | | | ||
Db     716 ndepipgsqagqiadtsnlfpvqthkstglpidpmvglpfdpksgnlvhpytnqtmsgls 775

Qy     828 TFESSSMNTPTSISSD-----------GMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
           :: |   :: :|          | |   :|:  :  | ||      |     :
Db     776 vsylaakn--ltvdtdetyglpidtltgypldpvslipfnpetgelfdpisdeimn-gti 832

Qy     877 SGIVTNPDSNES--SIVTSTVPTASTM-----SDSLSSTDGI----------SATSSDNV 919
           :|||:   ::||  |  :: :  |: |       |:  |:                |  :
Db     833 agivsgisasesllsqksalidpatnmvvgefggllnpatgvmipgflgpseqtqfspei 892

Qy     920 SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSES----ESKVTFTSN--GDN 973
              |:              |  ::::     :|| ||| ||    : |:   |    |
Db     893 edggi------------ippevaaanadkfklsippsvpesipekdqkidsiselmydi 939

Qy     974 QSGTHDSQ-----------------STSTEIEIVT------TSSTKVLPPVVSSNTDLT 1009
           :||    |                 | |: : ||      |:     ||     |
Db     940 esgrligqvskrpipgsiagdlnpimktptqtdsvtgkpidpttglpfnpp-----tghl 994

Qy    1010 SEPTNTREQPTTLS-----TTSNSITEDITTSQPTGD-NGDNTSSTNPVPTVATS----T 1059
             |||     ::  :      || |  |     | |:  |      : :|  :|:
Db     995 inptnnntmdssfagaykyavsngiktdnvyglpvgeitglpkdpgsdipfnsttgelvd 1054
```

```
Qy   1060 LASASEEDNKSGSHESASTSLKPSMGENSGL---TTSTEIEATTTSPTEAPSPAVSSGTD 1116
Db   1055 pstgkpinnstagivsgkpglppiedengnlfdpstnlpidgnnqlvnpetnstvsgsts 1114

Qy   1117 VTTEP--------------TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTS 1162
Db   1115 gttkpkpgipvngggvvpdeeakdq-----adkgkdglivpptnsin----kdpvtntqy 1165

Qy   1163 SLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
Db   1166 snttg----niinpe---tgkvipgslpgslnypsf------------ntpqqt--deit 1204

Qy   1223 SSSPSTNTFIASTYDGS-GSIIQHSTWL 1249
Db   1205 gkpvdtvtgl--pydpstgeiidpatkl 1230


  arch completed: October  3, 2002, 15:32:14
  b time: 366 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:26:33 ; Search time 16.32 Seconds
(without alignments)
1885.800 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	940	14.5	1537	1	US-08-325-267A-2	Sequence 2, Appli
2	516	7.9	894	3	US-08-362-525-22	Sequence 22, Appl
3	516	7.9	894	3	US-08-971-692-15	Sequence 15, Appl
4	488	7.5	1721	3	US-08-928-361B-6	Sequence 6, Appli
5	487	7.5	1721	3	US-08-700-651-5	Sequence 5, Appli
6	487	7.5	1837	3	US-08-928-361B-5	Sequence 5, Appli
7	468.5	7.2	862	1	US-08-325-267A-4	Sequence 4, Appli
8	404.5	6.2	2035	1	US-08-046-585-5	Sequence 5, Appli
9	404.5	6.2	2035	1	US-08-393-703-5	Sequence 5, Appli
10	404.5	6.2	2035	5	PCT-US93-11721-5	Sequence 5, Appli
11	350	5.4	650	3	US-08-362-525-2	Sequence 2, Appli
12	323	5.0	907	3	US-08-783-774-2	Sequence 2, Appli
13	323	5.0	907	5	PCT-US95-04611A-19	Sequence 19, Appl
14	322	5.0	1481	2	US-08-616-844-40	Sequence 40, Appl
15	322	5.0	1481	2	US-08-599-654-40	Sequence 40, Appl
16	322	5.0	1481	3	US-08-944-868A-40	Sequence 40, Appl
17	322	5.0	1481	3	US-08-944-423A-40	Sequence 40, Appl
18	322	5.0	1481	3	US-08-944-496-40	Sequence 40, Appl
19	310.5	4.8	249	3	US-08-700-651-15	Sequence 15, Appl
20	310.5	4.8	249	3	US-08-928-361B-20	Sequence 20, Appl
21	304.5	4.7	2409	6	5180808-2	Patent No. 5180808
22	304	4.7	1203	4	US-09-351-200-2	Sequence 2, Appli
23	303	4.7	216	3	US-08-928-361B-27	Sequence 27, Appl
24	301	4.6	216	3	US-08-928-361B-8	Sequence 8, Appli
25	297	4.6	750	4	US-09-165-239A-4	Sequence 4, Appli
26	296	4.6	1848	4	US-08-296-791-6	Sequence 6, Appli
27	296	4.6	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
28	277	4.3	3969	4	US-08-061-376-5	Sequence 5, Appli
29	276.5	4.3	933	3	US-08-293-728-2	Sequence 2, Appli
30	276.5	4.3	933	4	US-09-421-868-2	Sequence 2, Appli
31	269.5	4.1	1026	2	US-08-614-377A-7	Sequence 7, Appli
32	269.5	4.1	1026	4	US-09-142-648B-7	Sequence 7, Appli
33	268.5	4.1	1026	1	US-08-194-290-7	Sequence 7, Appli
34	268	4.1	1912	1	US-08-409-995-4	Sequence 4, Appli
35	268	4.1	1912	3	US-08-685-467-4	Sequence 4, Appli
36	267.5	4.1	2353	4	US-09-377-155-33	Sequence 33, Appl
37	267.5	4.1	2353	4	US-08-913-942-4	Sequence 4, Appli
38	267.5	4.1	2353	4	US-09-669-974-33	Sequence 33, Appl
39	266.5	4.1	2354	4	US-09-268-347-47	Sequence 47, Appl
40	266	4.1	2308	1	US-08-015-973-1	Sequence 1, Appli
41	266	4.1	2308	2	US-08-448-164-1	Sequence 1, Appli
42	266	4.1	2308	4	US-08-081-929-2	Sequence 2, Appli
43	263	4.0	1231	3	US-08-904-263A-4	Sequence 4, Appli
44	259.5	4.0	2411	4	US-09-268-347-36	Sequence 36, Appl
45	251.5	3.9	2628	2	US-08-570-311-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1537 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-08-325-267A-2

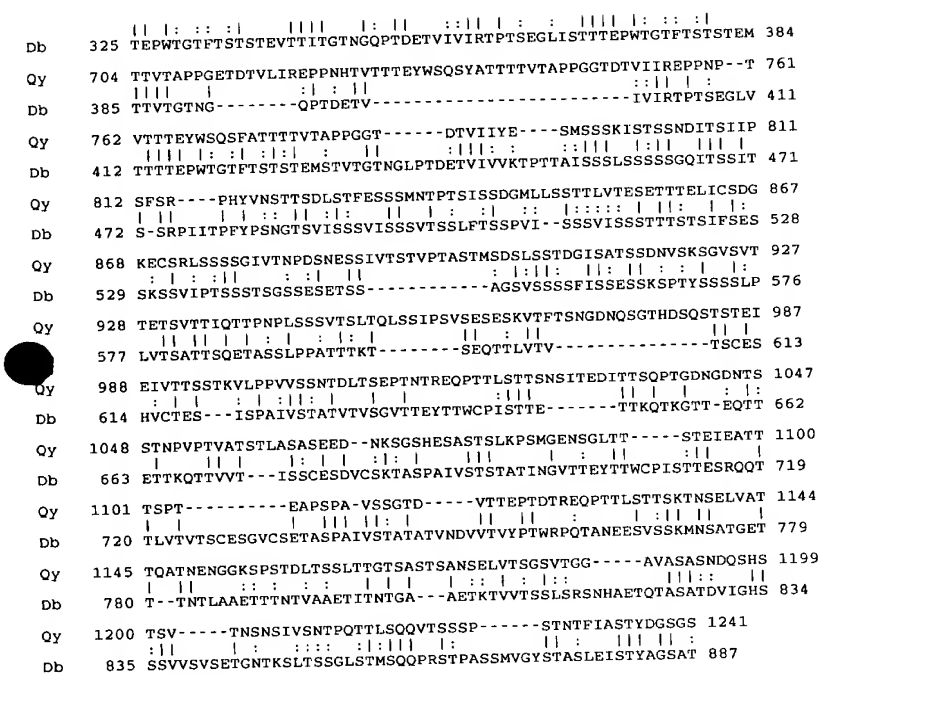
Query Match 14.5%; Score 940; DB 1; Length 1537;
Best Local Similarity 26.0%; Pred. No. 3.5e-46;
Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;

```
Qy       2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
Db      10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy      41 ----------------------------KGPGYPTW-----------NAVLGWSLD--G 58
Db      69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy      59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
Db     129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy      86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
Db     188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy     132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
Db     239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy     191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
Db     278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy     251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
Db     338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy     311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT---------- 355
Db     385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGL 438

Qy     356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
Db     439 PTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTEMTTVTGTNGLPTDET-IIVIRTPTT 497

Qy     399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
Db     498 ATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 556

Qy     448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
       557 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTELTTVTGTN 616

       495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
Db     617 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 676

Qy     543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
Db     677 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDT 736

Qy     592 YATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTVT--- 635
Db     737 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTN 796

Qy     636 GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
Db     797 GQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVIRTPT 856

Qy     687 NHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEYWSQS 735
Db     857 SEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGT 916

Qy     736 YATT----TTVTAPPG-GTD--TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPPGGT- 785
```

```
Db     917 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTITGTN 976

Qy     786 -----DTVIIYESMSSSKI------------STSSNDITSIIPSFSRP------------ 816
Db     977 GQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPT 1036

Qy     817 --HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICSDGKE 869
Db    1037 SEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQI 1096

Qy     870 CSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
Db    1097 TSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTSTSIF 1155

Qy     928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
Db    1156 SESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS----L 1205

Qy     988 EIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITEDITT 1035
Db    1206 PLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISPAIVS 1256

Qy    1036 SQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
Db    1257 TATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCSKTAS 1316

Qy    1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELVATT- 1145
Db    1317 PAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAIVSTA 1374

Qy    1146 ---------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSVTGGA 1188
Db    1375 TATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITNTGAA 1434

Qy    1189 ------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
Db    1435 ETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1494

Qy    1225 SP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db    1495 PASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536
```

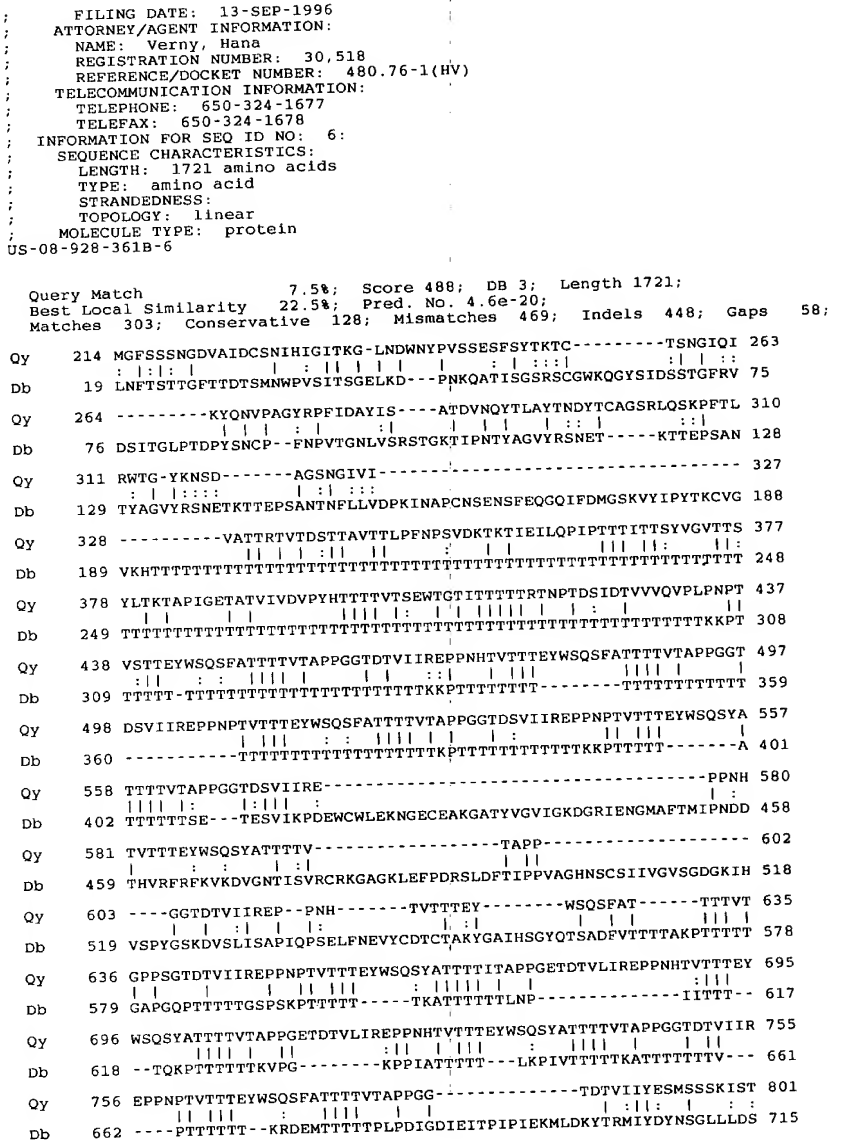
RESULT 2
US-08-362-525-22
; Sequence 22, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```
;       APPLICATION NUMBER:  EP 92202080.5
;       FILING DATE:  08-JUL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 92203899.7
;       FILING DATE:  14-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/EP93/01763
;       FILING DATE:  07-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOKULIS, PAUL N.
;       REGISTRATION NUMBER:  16,773
;       REFERENCE/DOCKET NUMBER:  213289/T7020(V)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;       TELEX:  6714627 CUSH
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  894 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-362-525-22


  Query Match             7.9%;  Score 516;  DB 3;  Length 894;
  Best Local Similarity   26.3%;  Pred. No. 4.5e-22;
  Matches  251;  Conservative  126;  Mismatches  310;  Indels  266;  Gaps   43;

Qy    437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
          | |   | :  :|: | : :  | ||  |     |  : : |
Db     53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy    478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
                  |||   |   ||                | | |:  | |         ::  ||
Db    113 CSNSQGIAYWSTDLFGFYTT--------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy    522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
          |   :    ||  :        | :|:| :      |     | ||:       |||
Db    158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy    581 TVTTTEYWSQSY------------ATTTTVTAPPGGT-----------------DTVI 609
          |   |   |                    :|| | | |
Db    205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy    610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
          : :| |:    | |||| |: :| :|    |||||     |:     ::|| | :  :  ||
 b    265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTT 324

Qy    657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
          || |: :: :|    ||||     |: ||   ::|| | :  :  |||| |: :: :|
Db    325 TEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEM 384

Qy    704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
          ||||   |        :| : ||                    ::||  | :
Db    385 TTVTGTNG--------QPTDETV-----------------------IVIRTPTSEGLV 411

Qy    762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
          |||| |: :| :|:|:|   :    ||      :|||: :    ::|||  |:||  ||| |
Db    412 TTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQITSSIT 471

Qy    812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
          | ||    | | :: || :|:    ||  | :  :|  ::  |::::: | ||:  | |:
Db    472 S-SRPIITPFYPSNGTSVISSSVISSSVTSSLFTSSPVI--SSSVISSSTTTSTSIFSES 528

Qy    868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
          : | : :||    : :|  ||                : |:||:  ||: || : : |  |:
Db    529 SKSSVIPTSSSTSGSSESETSS-----------AGSVSSSSFISSESSKSPTYSSSSLP 576

Qy    928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
          || || | | : |  : |: |         ||  : ||                  || |
Db    577 LVTSATTSQETASSLPPATTTKT--------SEQTTLVTV--------------TSCES 613

Qy    988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
          : | |   : | :||: |   |  |          :|||          || |  |   : |:
Db    614 HVCTES---ISPAIVSTATVTVSGVTTEYTTWCPISTTE-------TTKQTKGTT-EQTT 662

Qy   1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
            |    || |    |: | |  :|: |    |||     | :  ||      :||    |
Db    663 ETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESRQQT 719

Qy   1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
          |  |           |  ||| ||: |      ||  ||    :      | :|| ||    |
Db    720 TLVTVTSCESGVCSETASPAIVSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGET 779

Qy   1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
          |  ||    :: :  : :  | | |   | :: | : |::          |||::   ||
Db    780 T--TNTLAAETTTNTVAAETITNTGA---AETKTVVTSSLSRSNHAETQTASATDVIGHS 834

Qy   1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
            :||     | :   :::: :|:|||  |:          || :   ||| || :
Db    835 SSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGYSTASLEISTYAGSAT 887


RESULT   3
US-08-971-692-15
; Sequence 15, Application US/08971692
; Patent No. 6114147
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Immobilized proteins with specific binding
;    TITLE OF INVENTION:  capacities and their use in processes and products.
;    NUMBER OF SEQUENCES:  40
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/971,692
;      FILING DATE:
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-971-692-15


  Query Match             7.9%;  Score 516;  DB 3;  Length 894;
  Best Local Similarity   26.3%;  Pred. No. 4.5e-22;
  Matches  251;  Conservative  126;  Mismatches  310;  Indels  266;  Gaps   43;

Qy    437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
          | |   | :  :|: | : :  | ||  |     |  : : |
Db     53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy    478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
                  |||   |   ||                | | |:  | |         ::  ||
Db    113 CSNSQGIAYWSTDLFGFYTT--------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy    522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
          |   :    ||  :        | :|:| :      |     | ||:       |||
Db    158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy    581 TVTTTEYWSQSY------------ATTTTVTAPPGGT-----------------DTVI 609
          |   |   |                    :|| | | |
Db    205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy    610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
          : :| |:    | |||| |: :| :|    |||||     |:     ::|| | :  :  ||
Db    265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTT 324

Qy    657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
```



RESULT 4
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-6

Query Match 7.5%; Score 488; DB 3; Length 1721;
Best Local Similarity 22.5%; Pred. No. 4.6e-20;
Matches 303; Conservative 128; Mismatches 469; Indels 448; Gaps 58;



```
Qy    802 SSN----------DITSIIP---------------------SFSRPHYVNSTTSDLS 827
Db    716 NDEPIPGSQAGQIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLS 775

Qy    828 TFESSSMNTPTSISSD----------GMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
Db    776 VSYLAAKN--LTVDTDETYGLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMN-GTI 832

Qy    877 SGIVTNPDSNES--SIVTSTVPTASTM-----SDSLSSTDGI----------SATSSDNV 919
Db    833 AGIVSGISASESLLSQKSALIDPATNMVVGEFGGLLNPATGVMIPGFLGPSEQTQFSPEI 892

Qy    920 SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSES----ESKVTFTSN--GDN 973
Db    893 EDGGI------------IPPEVAAANADKFKLSIPPSVPESIPEKDQKIDSISELMYDI 939

Qy    974 QSGTHDSQ-----------------STSTEIEIVT------TSSTKVLPPVVSSNTDLT 1009
Db    940 ESGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVTGKPIDPTTGLPFNPP-----TGHL 994

Qy   1010 SEPTNTREQPTTLS-----TTSNSITEDITTSQPTGD-NGDNTSSTNPVPTVATS----T 1059
Db    995 INPTNNNTMDSSFAGAYKYAVSNGIKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGELVD 1054

Qy   1060 LASASEEDNKSGSHESASTSLKPSMGENSGL---TTSTEIEATTTSPTEAPSPAVSSGTD 1116
Db   1055 PSTGKPINNSTAGIVSGKPGLPPIEDENGNLFDPSTNLPIDGNNQLVNPETNSTVSGSTS 1114

Qy   1117 VTTEP--------------TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTS 1162
Db   1115 GTTKPKPGIPVNGGGVVPDEEAKDQ-----ADKGKDGLIVPPTNSIN----KDPVTNTQY 1165

Qy   1163 SLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
Db   1166 SNTTG----NIINPE---TGKVIPGSLPGSLNYPSF------------NTPQQT--DEIT 1204

Qy   1223 SSSPSTNTFIASTYDGS-GSIIQHSTWL 1249
Db   1205 GKPVDTVTGL--PYDPSTGEIIDPATKL 1230
```

RESULT 5
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match 7.5%; Score 487; DB 3; Length 1721;
Best Local Similarity 22.3%; Pred. No. 5.2e-20;
Matches 293; Conservative 113; Mismatches 411; Indels 498; Gaps 52;

```
Qy    297 TCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTT-----LPFNPSVD 351
Db     52 TISGSRS---------CGWK------QGYSIDSSTGFRVDSITGLPTDPYSNCPFNPVTG 96

Qy    352 KTKTIEILQPIPTT---------TITT-----SYVGVTTSYLTKTA-PIGETATVIVD-- 394
Db     97 NLVSRSTGKTIPNTYAGVYRSNETKTTEPSANTYAGVYRSNETKTTEPSANTNFLLVDPK 156

Qy    395 ------------------------VPY-------HTTTTVTSEWTGTITTTTTRTNPTDS 423
Db    157 INAPCNSENSFEQGQIFDMGSKVYIPYTKCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTT 216

Qy    424 IDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQS 483
Db    217 TTT----------TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 266

Qy    484 FATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPN 543
Db    267 TTTTTTTT----------TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTT 309

Qy    544 PTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPG 603
Db    310 TTTTTTTTTTTTTTTTTTTT----TTTTTTKKPTTTTTTT-------TTTTTTTT--- 355

Qy    604 GTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQS 663
Db    356 -TTTT--------TTTTTTTTTTTTTTTTTTKPTTTTTTTTTTTTKKPTTTT------ 400

Qy    664 YATTTTIT---------------------------------------------------- 671
Db    401 -ATTTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMIPNDDT 459

Qy    672 ---------------------------------------------------------- 671
Db    460 HVRFRFKVKDVGNTISVRCRKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVSGDGKIHV 519

Qy    672 APPGETDTVLIREPPNHTVTTTEYWSQSYA--------------------------TTT 704
Db    520 SPYGSKDVSLISAP----IQPSELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTTAKPTT 575

Qy    705 TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTT 764
Db    576 TTTGAPGQPTTTTTGSPSKPTTTTT-----TKATTTTTTLNPIITTTT---QKPTTTTTT 627

Qy    765 TEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTS 824
Db    628 KVPGKPPIATTTTTLKPIVTTTT---------TKATTTT---TTTVP---------TTTT 666

Qy    825 DLSTFESSSMNTP---------TSISSDGMLLSSTTLVTESETTTELICSD----GKECS 871
Db    667 TTKRDEMTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEPIPGSQAG 726

Qy    872 RLS-----------SSSGIVTNP------DSNESSIV-TSTVPTASTMSDSLSSTDGISA 913
Db    727 QIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNLTV 786

Qy    914 TSSDNVSKSGVSVTTETS-----VTTIQTTP------NPLSSSVTSLTQLSSIPSVSESE 962
Db    787 DTDETY---GLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSGISASE 843

Qy    963 S----KVTFTSNGDNQ--------------------SGTHDSQSTSTEIEIVTTSSTKVL 998
Db    844 SLLSQKSALIDPATNMVVGEFGGLLNPATGVMIPGFLGPSEQTQFSPEIE-----DGGII 898

Qy    999 PP-VVSSNTD--LTSEPTNTREQPTTLSTTSNSITE---DITTSQPTGD------NGDNT 1046
Db    899 PPEVAAANADKFKLSIPPSVPESIPEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIA 958

Qy   1047 SSTNPVPTVATST---------------------LASASEEDNKSGSHESA-----STS 1079
Db    959 GDLNPIMKTPTQTDSVTGKPIDPTTGLPFNPPTGHLINPTNNNTMDSSFAGAYKYAVSNG 1018

Qy   1080 LKPS------MGENSGL-----------TTSTEIEATTTSPTEAPSPAVSSGT------ 1115
Db    1019 GIKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGELVDPSTGKPINNSTAGIVSGKPGLP 1078
```

```
Db   1019 IKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGELVDPSTGKPINNSTAGIVSGKPGLPPI 1078

Qy   1116 --------DVTTE-PTDTREQ---PTTLSTTSKTNSELVATTQATNENGG---------- 1153
                  | :|  | |   |   | | || | : |           |||
Db   1079 EDENGNLFDPSTNLPIDGNNQLVNPETNSTVSGSTSGTTKPKPGIPVNGGGVVPDEEAKD 1138

Qy   1154 ------------------KSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASND 1195
                             | | |:   | |||       | |   :| |  |::  : |
Db   1139 QADKGKDGLIVPPTNSINKDPVTNTQYSNTTG----NIINPE---TGKVIPGSLPGSLNY 1191

Qy   1196 QSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGS-GSIIQHSTWL 1249
           |             |||| |   ::|     | | :   || | | ||  :| |
Db   1192 PSF------------NTPQQT--DEITGKPVDTVTGL--PYDPSTGEIIDPATKL 1230


RESULT   6
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
     APPLICANT:  Petersen, Carolyn
     TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
     TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1837 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-5


  Query Match             7.5%;  Score 487;  DB 3;  Length 1837;
  Best Local Similarity   22.8%;  Pred. No. 5.7e-20;
  Matches  316;  Conservative  146;  Mismatches  487;  Indels  440;  Gaps   61;

Qy    136 VGGTGSSTDLEDSKCFTAGTNTVTFN-DGDKDISIDVEFEKST---VDP-SAYLY--ASR 188
          :  ||        |   | | :  ::   : |:|  |  | :|   ||| |  |:  ::
Db     69 IDSTGKKFSPYTGKHADASTTSSAYSAPFELDVS-GVPIEPNTRRMVDPVSLMLFDNSTG 127

Qy    189 VM--PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
          ||  |: | :    :|          : |:|: |       |  : || |
Db    128 VMYDPNTNSILEGSIAGIRSESCIVSELNFTSTTGFTTDTSMNWPVSITSG--------- 178
```

```
Qy    247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSK 306
                      ::|  |  |                                | :|||
Db    179 --------------ELKDPNKQA-------------------------TISGSRS--- 194

Qy    307 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTT-----LPFNPSVDKTKTIEILQP 361
              |:|      |  | ::|   || | : |     ||||      |  ::
Db    195 ------CGWK------QGYSIDSSTGFRVDSITGLPTDPYPNCPFNP-----VTGNLVSR 237

Qy    362 IPTTTITTSYVGVTTSYLTKTA-PIGETATVIVD------------------------- 394
            ||  :| ||  |  |||  |   |  ::||
Db    238 STGKTIPNTYAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSENSFEQVQIFDMGSKVY 297

Qy    395 VPY-------HTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQS 447
          :||       ||||| |:  | | ||||| |  | :  |           | :||   : :
Db    298 IPYTKCVGVKHTTTTTTT-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 356

Qy    448 FATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPN 507
            |||| |     | |   ::|   | |||   : :  |||| |
Db    357 TTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTT---------------T 401

Qy    508 PTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPG 567
            | |||   : :  |||| |  |  | :          | |||   : :  |||| |  |
Db    402 TTTTTTTTTTTTTTTTTTTKKPTTTTTT----TTTTTTTTTTTTTTTTTTTTTKKPT 457

Qy    568 GTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQS 627
            | :          | |||   : :  |||| ||       |   ::|   | |||   ::
Db    458 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA------TTTTKKPTTTTTTTTTTTKK 511

Qy    628 FATTTTVTGPPSGTDTVIIREP------------------------------------- 649
            |||| |   : ::|  : :|
Db    512 PTTTTTATTTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMI 571

Qy    650 PN-PTVTTTEYWSQSYATTTTI-----------------TAPP---------------- 674
            ||  |      :  :    | ::                 | ||
Db    572 PNDDTHVRFRFKVKDVGNTISVRCGKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVSGG 631

Qy    675 --------GETDTVLIREPPNHTVTTTEYWSQSYATT------------------TTVT 707
                  |  |  ||  |    :   | ::: |  |                   || |
Db    632 GKIHVSPYGSKDVSLISAP----IQPCELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTT 687

Qy    708 APPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEY 767
            | |  | |    :|     |||   | |  |||| |      | | |:    || :|||
Db    688 AKPTTTTTGAPGQP-----TTTTTGSPSKPTTTTTT--KATTTTTIL----NPIITTT-- 734

Qy    768 WSQSFATTTTVTA---PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTS 824
            :|   ||||      ||  | |  :   ::::    ::   |:          | :||:
Db    735 -TQKPTTTTTTKVPGKPPIATTTTTLKPIVTTTTTKATTTTTTT----------VPTTTT 783

Qy    825 DLSTFESSSMNTP---------TSISSDGMLLSSTTLVTESETTTELICSD----GKECS 871
            | ::  ||          | |  : ||   | :: : :   |  :|    | :
Db    784 TTKRDEMTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEPIPGSQAG 843

Qy    872 RLS-----------SSSGIVTNP------DSNESSIV-TSTVPTASTMSDSLSSTDGISA 913
            :::           |:|:  :|      |    ::|   |  | | :| |  :   ::
Db    844 QIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNLTV 903

Qy    914 TSSDNVSKSGVSVTTETS-----VTTIQTTP------NPLSSSVTSLTQLSSIPSVSESE 962
            : :     |: : | |       |: |   |       :|:|  : : |    :  :| ||
Db    904 DTDETY---GLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSGISASE 960

Qy    963 SKVTFTSN----------GD--------------NQSGTHDSQSTSTEIEIVTTSSTKVL 998
            | ::  |          |:                |   :    | |||       ::
Db    961 SLLSQKSAPIDPATNMVVGEFGGLLNPATGVMIPGSLGPSEQTPFSPEIE-----DGGII 1015

Qy    999 PP-VVSSNTD--LTSEPTNTREQPTTLSTTSNSITE---DITTSQPTGD------NGDNT 1046
            || | ::| |    | | :  |          :||:|   || : :  |         |
Db   1016 PPEVAAANADKFKLSIPPSVPESIPEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIA 1075

Qy   1047 SSTNPVPTVATSTLASASEEDNKSGSHESASTSL--KPSMGENSGLTTSTEIEATTTSPT 1104
             ||:    | |       |: :|     :| |   |  |    |   |       :|
Db   1076 GDLNPIMKTPTQT-------DSVTGKPIDPTTGLPFNPPTGH---LINPTNNNTMDSSFA 1125
```

```
Qy   1105 EAPSPAVSSG--TD-----VTTEPTDTREQPTTLSTTSKTNSELV--ATTQATN------ 1149
           |   |||:|  ||           | |    : | :      : |  |||  :| :  |
Db   1126 GAYKYAVSNGIKTDNVYGLPVDEITGLPKDPVSDIPFNSTTGELVDPSTGKPINNYTAGI 1185

Qy   1150 ------------ENGGK-SPSTDL------------TSSLTTG-TSASTSANSELVTSGS 1183
                      |||    ||| |             |:|  :| || ||      :  :|
Db   1186 VSGKRGLPPIEDENGNLFDPSTKLPIDGNNQLVNPETNSTVSGSTSGSTKPKPGIPVNG- 1244

Qy   1184 VTGGAVA-SASNDQSHSTS----VTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDG 1238
            | |  |    : ||:         |  :|||                :  |  |||   ::|
Db   1245 --GGVVPDEEAKDQADKGKDGLIVPPTNSI--------------NKDPVTNTQYSNT--- 1285

Qy   1239 SGSIIQHST 1247
           :|:||   |
Db   1286 TGNIINPET 1294
```

RESULT 7
S-08-325-267A-4
Sequence 4, Application US/08325267A
Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 862 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-267A-4

```
  Query Match            7.2%;  Score 468.5;  DB 1;  Length 862;
  Best Local Similarity   26.3%;  Pred. No. 2.3e-19;
  Matches  231;  Conservative  120;  Mismatches  297;  Indels  229;  Gaps   40;

Qy    545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTT------------------ 585
          | :    | :  ||: | : :   | ||   |      |  : : |
Db     53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy    586 -------EYWSQS----YATTTTVT--------APPGGT---------DTVII------- 610
                 |||     | | | ||           | |:          |: |:
Db    113 CSNSQGIAYWSTDLFGFYTTPTNVTLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATA 172

Qy    611 ------REPP----NHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIRE----PPNPTVTT 656
                :: ||    | |:   :  |  |              ||:    || :      |   :
Db    173 FNCCAQQQPPITSTNFTIDGIKPWGGSL--------PPNIEGTVYMYAGYYYPMKVVYSN 224

Qy    657 TEYWSQSYATTTTITAPPGET-----------------DTVLIREPPNH----TVTTTE 694
             |          ::| | | |                     : :| |:     | ||||
Db    225 AVSWG---TLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCTVPDPSNYAVSTTTTTTE 281

Qy    695 YWSQSYATT----TTVTAP---PGETDTVLIREPPNHT--VTTTEYWSQSYATT----TT 741
          |: :: :|      ||||       :  ::|| |   :  :||||  |: :: :|       ||
Db    282 PWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWTGTFTSTSTEMTT 341

Qy    742 VTAPPG-GTD--TVIIREPPNP--TVTTTEYWSQSFATTTTVTAPPGGT------DTVII 790
          ||     |  ||     ::|| | :     |||| |: :| :|:|    :    ||      :|||:
Db    342 VTGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIV 401

Qy    791 YE----SMSSSKISTSSNDITSIIPSFSR----PHYVNSTTSDLSTFESSSMNTPTSISS 842
            :     ::|||   |:||   |||  |  |  ||        | |  :: || |:  || |  : :|
Db    402 VKTPTTAISSSLSSSSSGQITSSITS-SRPIITPFYPSNGTSVISSSVISSSVTSSLFTS 460

Qy    843 DGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMS 902
           ::  |:::::  | ||:  | |:  :  | :  :||    :  :| ||                  :
Db    461 SPVI--SSSVISSSTTTSTSIFSESSKSSVIPTSSSTSGSSESETSS-----------A 506

Qy    903 DSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESE 962
           |:||:  ||: || : : |   |:    || || |  | : |   : |: |          ||
Db    507 GSVSSSSFISSESSKSPTYSSSSLPLVTSATTSQETASSLPPATTTKT--------SEQT 558

Qy    963 SKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTL 1022
           : ||                 || |   : | |   : | :||: |    |   |             :
Db    559 TLVTV--------------TSCESHVCTES---ISPAIVSTATVTVSGVTTEYTTWCPI 600

Qy   1023 STTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEED--NKSGSHESASTSL 1080
           |||          || |  |    : |: |     || |      |: | |   :|: |    |||
Db    601 STTE-------TTKQTKGTT-EQTTETTKQTTVVT---ISSCESDVCSKTASPAIVSTST 649

Qy   1081 KPSMGENSGLTT-----STEIEATTTSPT----------EAPSPA-VSSGTD-----VTT 1119
             |  :  ||        :||     ||  |              |  |||  ||: |        ||
Db    650 ATINGVTTEYTTWCPISTTESRQQTTLVTVTSCESGVCSETASPAIVSTATATVNDVVTV 709

Qy   1120 EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
            ||    :       | :|| ||      ||  ||      :: :    : :      | | |    | :: |
Db    710 YPTWRPQTANEESVSSKMNSATGETT--TNTLAAETTTNTVAAETITNTGA---AETKTV 764

Qy   1180 TSGSVTGG-----AVASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSSSP--- 1226
           : |::             |||::     ||:||      | :      :::: :|:|||  |:
Db    765 VTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSM 824

Qy   1227 ---STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
               || :      ||| || : :    : |    |  | | |
Db    825 VGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 861
```

RESULT 8
US-08-046-585-5
; Sequence 5, Application US/08046585
; Patent No. 5453362
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/046,585
;      FILING DATE:  12-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57503-1/RAO
     TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-046-585-5
```

```
  Query Match              6.2%;  Score 404.5;  DB 1;  Length 2035;
  Best Local Similarity   21.9%;  Pred. No. 3.7e-15;
  Matches  319;  Conservative  165;  Mismatches  506;  Indels  467;  Gaps   69;

Qy      14 IASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAV-LGWSLDGTSA-NPGDTFTLNM  71
           | |:  ::|:     :|  : ||     ||  :|  | ::  | |  ||    | :
Db     537 IGSSPQMSGM----AALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG---TTTL 589

Qy      72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLP  131
           |   |  :|   |                |        |    |  :: |   |  | | |
Db     590 PATVKVASSPVMV---------------SNPATRMLKTAAAQVGTSVSS---ATNTSTRP 631

Qy     132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP  191
           |   |  :|: |  : ::  |     ||     | |::          | :    | ::
Db     632 I-ITVHKSGTVTVAQQAQVVTTVVGGVT-----KTITL-------VKSPISVPGGSALIS 678

Qy     192 SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS  251
           :| || ::    | :   ||   |  :| | |          | |||
Db     679 NLGKVMSVV---QTKPVQTSAVTG-QASTGPVT------QIIQTKG-------------- 714

Qy     252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR  311
                       :|||    |   ::: |    |:  |   : ||    :||  |
Db     715 ---------------PLPAG---TILKLVTSADGKPTTIITTTQASGAG----TKPTIL- 751

Qy     312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDK------TKTIEILQPIP--  363
                     ||  |: : |   :|| : |:| :  : :       | :  |  ||
Db     752 ------------GISSVSPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITII 799

Qy     364 TTTITTSYVGVTTSYLT------------------KTAPIGETATVIVDVPYHTTTTVT  404
           || : ||  |   :|                   | || |:  |::  ||      ||
Db     800 TTKVMTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAP-GQPGTILRTVPMGGVRLVT 858

Qy     405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV  464
                 | :   |  :: |:||:      |: |         | :|  |  ||  |
Db     859 -------PVTVSAVKP--AVTTLVVKGTTGVTTLGTVT------GTVSTSLAGAGGHST- 902

Qy     465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT  524
                       | | ||  |      || : :  :  |||  |        | |
Db     903 ---------------SASLATPITTL----GTIATLSSQVINPTAITVS------AAQT 936
```

```
Qy     525 TVTAPPGGTDSVIIREP-PNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTV- 582
           |:||  | |   |  :|   ||              | :||| |      :   | | :
Db     937 TLTAAGGLTTPTITMQPVSQPT-----------QVTLITAPSG------VEAQPVHDLP 978

Qy     583 ------TTTEYWSQSYATTTTVTAPPG----GTDTVIIREPP--NHTVTTTEYWSQSFAT 630
                 |||   | || |   :  |    || |::  ||  |    ||   : ||
Db     979 VSILASPTTE---QPTATVTIADSGQGDVQPGTVTLVCSNPPCETHETGTT-----NTAT 1030

Qy     631 TTTVT---GPPSG----------------TDTV------IIREPPNPTVTTTEYWSQSY 664
           || |    | |                   | ||      ::|   ||   | |  : :
Db    1031 TTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNT 1090

Qy     665 ATTTT--------ITAPPGET---------DTVLIREPPNH----------------T 689
           ||| |          : || ||          | :    ||                :
Db    1091 ATTATSNMAGQHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAAGTPAVIRIS 1150

Qy     690 VTTTEYWS----------QSYATTTTVT-----AP--------------PGETDTVLIR 719
           | |   :           |: ||:||:|     ||              ||      ::
Db    1151 VATGALEAAQGSKSQCQTRQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210

Qy     720 EPP--------------------NHTVTTTEYWSQSY------------------ATTT 740
             |                    :| |:|     |                    :||
Db    1211 LAPLSSKVRLSSPSIKDLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPSTTV 1270

Qy     741 TVTA-----PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS 795
           ||||     |  | | :   ||  |          ::  | |  |:  |     |
Db    1271 TVTALEALLCPSATVTQVCSNPPCET-------HETGTTNTATTSNAGSAQRVCSNPPCE 1323

Qy     796 SSKISTSSNDITSII----------PSFSRPHYVNSTTSDLSTFE-SSSMNTPTSISS 842
           : :  |:   |:                |   ||   : |||  :|   |     | : ||
Db    1324 THETGTTHTATTATSNGGTGQPEGGQQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSS 1383

Qy     843 -----DGMLLSSTTLVTESETT-------TELICSDGKECSRLSSSSGIVTNP--DSNES 888
                |: :::   ||    |        |: :||                ||  :::|
Db    1384 HRTVESGLEVAAAPSVTPQAGTALLAPFPTQRVCS---------------NPPCETHE- 1426

Qy     889 SIVTSTVPTASTMSDSLSSTDGISATSSD-----NVSKSGVSVTTETSVTTIQTTPNPLS 943
             |  ||:|:: ::||        :||    :     |::|: :::||  |  : |:
Db    1427 ---TGTTHTATTVTSNMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITT--TVSSTLT 1481

Qy     944 SSVTSLTQLSSIPSVSESESKVTFTSNGDNQS--GTHDSQSTSTEI-----EIVTTSSTK 996
           :||::|| : :|  |    :    | |  |         || || :     |::: | |
Db    1482 RAVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSASTALMGESAEVLSASQTP 1541

Qy     997 VLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVA 1056
           ||  |  ::  | ||:: :|       :: |       ||           | || :
Db    1542 ELPAAVDLSS--TGEPSSGQE-------SAGSAVVATVVVQP-----------PPPTQS 1580

Qy    1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS 1113
                |  :|   : ::|:     | | |   |   :| : |  |   :  || : |: :
Db    1581 EVDQLSLPQELMAEAQAGTTTLMVTGLTP---EELAVTAAAEAAAQAAATEEAQALAIQA 1637

Qy    1114 GTDV-------TTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTT 1166
                     | || || |   |:      | :||       :  | :  :| :   ||
Db    1638 VLQAAQQAVMGTGEPMDTSEAAATV-----TQAEL----GHLSAEGQEGQATTIPIVLTQ 1688

Qy    1167 GTSASTSANSELVTSGS--------VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216
              |:     :|  : :           | | : || :  ::    | :    | |
Db    1689 QELAALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746

Qy    1217 LSQQVTSSSPSTNTFIA 1233
           |   | |   :|||:|
Db    1747 LPSTATESLAPSNTFVA 1763
```

```
RESULT   9
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
```

```
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,703
;      FILING DATE:  24-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57503-2/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-393-703-5
```

Query Match 6.2%; Score 404.5; DB 1; Length 2035;
Best Local Similarity 21.9%; Pred. No. 3.7e-15;
Matches 319; Conservative 165; Mismatches 506; Indels 467; Gaps 69;

```
Qy      14 IASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAV-LGWSLDGTSA-NPGDTFTLNM  71
           | |:  ::|:     :|  : ||          ||  :|  | ::  | |  ||   |  :
Db     537 IGSSPQMSGM----AALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG---TTTL 589

Qy      72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLP 131
           |   |  :|   |              |     |  :: |   |  | | |
Db     590 PATVKVASSPVMV--------------SNPATRMLKTAAAQVGTSVSS---ATNTSTRP 631

Qy     132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP 191
           |   |  :|: |  : ::  |     ||     | |::          | :    | ::
Db     632 I-ITVHKSGTVTVAQQAQVVTTVVGGVT-----KTITL-------VKSPISVPGGSALIS 678

Qy     192 SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 251
           :| || ::    | :   ||   |  :| | |        |  |||
Db     679 NLGKVMSVV---QTKPVQTSAVTG-QASTGPVT------QIIQTKG------------- 714

Qy     252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
                  :|||   |  ::: |   |:  |   : ||    :||  |
Db     715 --------------PLPAG---TILKLVTSADGKPTTIITTTQASGAG----TKPTIL- 751

Qy     312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDK------TKTIEILQPIP-- 363
                    ||  |: : |   :|| : |:| :  :  : :       | :  |  ||
Db     752 -----------GISSVSPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITII 799

Qy     364 TTTITTSYVGVTTSYLT------------------KTAPIGETATVIVDVPYHTTTTVT 404
           || : ||  |     :|                   | || |:  |::  ||      ||
Db     800 TTKVMTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAP-GQPGTILRTVPMGGVRLVT 858

Qy     405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV 464
                 | :   |  :: |:||:      |: |         | :|  |  ||  |
Db     859 -------PVTVSAVKP--AVTTLVVKGTTGVTTLGTVT------GTVSTSLAGAGGHST- 902
```

```
Qy     465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT 524
                         | | ||  |      || : :  :  |||  |        |  |
Db     903 ---------------SASLATPITTL----GTIATLSSQVINPTAITVS------AAQT 936

Qy     525 TVTAPPGGTDSVIIREP-PNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTV- 582
           |:||  | |    |  :|   ||                | :||| |      :   | | :
Db     937 TLTAAGGLTTPTITMQPVSQPT-----------QVTLITAPSG------VEAQPVHDLP 978

Qy     583 ------TTTEYWSQSYATTTTVTAPPG----GTDTVIIREPP--NHTVTTTEYWSQSFAT 630
                 |||    |  || |   :  |    || |::   ||   |   ||      : ||
Db     979 VSILASPTTE---QPTATVTIADSGQGDVQPGTVTLVCSNPPCETHETGTT-----NTAT 1030

Qy     631 TTTVT---GPPSG----------------TDTV------IIREPPNPTVTTTEYWSQSY 664
           || |    | |                   | ||      ::|   ||   | |  : :
Db    1031 TTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNT 1090

Qy     665 ATTTT--------ITAPPGET---------DTVLIREPPNH-----------------T 689
           |||  |         : || ||          | :    ||                 :
Db    1091 ATTATSNMAGQHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAAGTPAVIRIS 1150

Qy     690 VTTTEYWS----------QSYATTTTVT-----AP--------------PGETDTVLIR 719
           | |    :           |: ||:||:|     ||              ||      ::
Db    1151 VATGALEAAQGSKSQCQTRQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210

Qy     720 EPP--------------------NHTVTTTEYWSQSY------------------ATTT 740
             |                    :| |:|      |                    :||
Db    1211 LAPLSSKVRLSSPSIKDLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPSTTV 1270

Qy     741 TVTA-----PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS 795
           ||||     | | : ||  |          ::   | |  |:  |    |     | : ||
Db    1271 TVTALEALLCPSATVTQVCSNPPCET-------HETGTTNTATTSNAGSAQRVCSNPPCE 1323

Qy     796 SSKISTSSNDITSII------------PSFSRPHYVNSTTSDLSTFE-SSSMNTPTSISS 842
           : :  |:    |:              |   ||   : |||  :|   |     | : ||
Db    1324 THETGTTHTATTATSNGGTGQPEGGQQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSS 1383

Qy     843 -----DGMLLSSTTLVTESETT-------TELICSDGKECSRLSSSSGIVTNP--DSNES 888
                |: :::   ||    |       |: :||                ||  :::|
Db    1384 HRTVESGLEVAAAPSVTPQAGTALLAPFPTQRVCS----------------NPPCETHE- 1426

Qy     889 SIVTSTVPTASTMSDSLSSTDGISATSSD-----NVSKSGVSVTTETSVTTIQTTPNPLS 943
           | |  ||:|:: ::||       :||     :     |::|: :::||  |  : |:
Db    1427 ---TGTTHTATTVTSNMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITT--TVSSTLT 1481

Qy     944 SSVTSLTQLSSIPSVSESESKVTFTSNGDNQS--GTHDSQSTSTEI-----EIVTTSSTK 996
           :|||::|| : :|  |    :    | |  |         || || :     |::: | |
Db    1482 RAVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSASTALMGESAEVLSASQTP 1541

Qy     997 VLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVA 1056
           ||  |  ::  | ||:: :|       :: |       ||            | || :
Db    1542 ELPAAVDLSS--TGEPSSGQE-------SAGSAVVATVVVQP------------PPPTQS 1580

Qy    1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS 1113
              |  :|   : ::|:     | | |   |  :| : |  |   :  || : |: :
Db    1581 EVDQLSLPQELMAEAQAGTTTLMVTGLTP---EELAVTAAAEAAAQAAATEEAQALAIQA 1637

Qy    1114 GTDV-------TTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTT 1166
                      | || || |   |:     | :||       :  | :  :| :   ||
Db    1638 VLQAAQQAVMGTGEPMDTSEAAATV-----TQAEL----GHLSAEGQEGQATTIPIVLTQ 1688

Qy    1167 GTSASTSANSELVTSGS--------VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216
              |:     :|  : :           | | : || :  ::    | :    | |
Db    1689 QELAALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746

Qy    1217 LSQQVTSSSPSTNTFIA 1233
           |    | |   :|||:|
Db    1747 LPSTATESLAPSNTFVA 1763
```

```
RESULT 10
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11721
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  FP-57503-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-11721-5


  Query Match             6.2%;  Score 404.5;  DB 5;  Length 2035;
  Best Local Similarity   21.9%;  Pred. No. 3.7e-15;
  Matches  319;  Conservative  165;  Mismatches  506;  Indels  467;  Gaps   69;

Qy      14 IASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAV-LGWSLDGTSA-NPGDTFTLNM 71
           | |:  ::|:     :|  : ||          ||  :|  | ::  | |  ||   |  :
Db     537 IGSSPQMSGM----AALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG---TTTL 589

Qy      72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLP 131
           |   |  :|   |              |       |    | :: |   |  | | |
Db     590 PATVKVASSPVMV---------------SNPATRMLKTAAAQVGTSVSS---ATNTSTRP 631

Qy     132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP 191
           |   |  :|: |  : ::  |      ||      | |::           |  :      | ::
Db     632 I-ITVHKSGTVTVAQQAQVVTTVVGGVT-----KTITL-------VKSPISVPGGSALIS 678

Qy     192 SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 251
           :| || ::    | :   ||   |  :| | |         |  |||
Db     679 NLGKVMSVV---QTKPVQTSAVTG-QASTGPVT------QIIQTKG-------------- 714

Qy     252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
                       :|||    |   ::: |    |:  |   : ||    :||  |
Db     715 ---------------PLPAG---TILKLVTSADGKPTTIITTTQASGAG----TKPTIL- 751

Qy     312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDK------TKTIEILQPIP-- 363
                  ||  |: : |   :|| : |:| :  : :  :       | :  |  ||
Db     752 ------------GISSVSPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITII 799

Qy     364 TTTITTSYVGVTTSYLT------------------KTAPIGETATVIVDVPYHTTTTVT 404
           || : ||  |     :|                  | || |:  |::  ||      ||
Db     800 TTKVMTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAP-GQPGTILRTVPMGGVRLVT 858

Qy     405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV 464
                 | :   |  :: |:||:      |: |         | :|  |  ||  |
Db     859 -------PVTVSAVKP--AVTTLVVKGTTGVTTLGTVT------GTVSTSLAGAGGHST- 902

Qy     465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT 524
                         | | ||  |      || : :  :  |||  |         | |
Db     903 ---------------SASLATPITTL----GTIATLSSQVINPTAITVS------AAQT 936

Qy     525 TVTAPPGGTDSVIIREP-PNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTV- 582
           |:||  | |   |  :|   ||              | :||| |      :   | | :
Db     937 TLTAAGGLTTPTITMQPVSQPT------------QVTLITAPSG------VEAQPVHDLP 978

Qy     583 ------TTTEYWSQSYATTTTVTAPPG----GTDTVIIREPP--NHTVTTTEYWSQSFAT 630
                  |||   |  || |   :  |    || |::   ||   |   ||      : ||
Db     979 VSILASPTTE---QPTATVTIADSGQGDVQPGTVTLVCSNPPCETHETGTT-----NTAT 1030

Qy     631 TTTVT---GPPSG-----------------TDTV------IIREPPNPTVTTTEYWSQSY 664
           || |    | |                    | ||      ::|   ||   | |  : :
Db    1031 TTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNT 1090

Qy     665 ATTTT--------ITAPPGET---------DTVLIREPPNH-----------------T 689
           ||| |         : || ||           | :      ||                :
Db    1091 ATTATSNMAGQHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAAGTPAVIRIS 1150

Qy     690 VTTTEYWS----------QSYATTTTVT-----AP-------------PGETDTVLIR 719
           | |   :            |: ||:||:|     ||             ||    ::
Db    1151 VATGALEAAQGSKSQCQTRQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210

Qy     720 EPP--------------------NHTVTTTEYWSQSY------------------ATTT 740
             |                     :| |:|      |                    :||
Db    1211 LAPLSSKVRLSSPSIKDLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPSTTV 1270

Qy     741 TVTA-----PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS 795
           ||||      |  | | :   ||  |        ::  | |  |:  |    |
Db    1271 TVTALEALLCPSATVTQVCSNPPCET-------HETGTTNTATTSNAGSAQRVCSNPPCE 1323

Qy     796 SSKISTSSNDITSII-----------PSFSRPHYVNSTTSDLSTFE-SSSMNTPTSISS 842
           : :  |:    |:               |   ||   : |||  :|   |      | : ||
Db    1324 THETGTTHTATTATSNGGTGQPEGGQQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSS 1383

Qy     843 -----DGMLLSSTTLVTESETT-------TELICSDGKECSRLSSSSGIVTNP--DSNES 888
                |: :::   ||    |       |: :||                ||  :::|
Db    1384 HRTVESGLEVAAAPSVTPQAGTALLAPFPTQRVCS---------------NPPCETHE- 1426

Qy     889 SIVTSTVPTASTMSDSLSSTDGISATSSD-----NVSKSGVSVTTETSVTTIQTTPNPLS 943
             |  ||:|:: ::||       :||    :     |::|: :::||  |  : |:
Db    1427 ---TGTTHTATTVTSNMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITT--TVSSTLT 1481

Qy     944 SSVTSLTQLSSIPSVSESESKVTFTSNGDNQS--GTHDSQSTSTEI-----EIVTTSSTK 996
           :||::|| : :|  |   :    | |  |       || || :     |::: | |
Db    1482 RAVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSASTALMGESAEVLSASQTP 1541

Qy     997 VLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVA 1056
           ||  |  ::  | ||:: :|       :: |        ||            | || :
Db    1542 ELPAAVDLSS--TGEPSSGQE-------SAGSAVVATVVVQP------------PPPTQS 1580

Qy     1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS 1113
              |  :|   : ::|:     | | |   |   :| : |  |   :  || : |: :
Db    1581 EVDQLSLPQELMAEAQAGTTTLMVTGLTP---EELAVTAAAEAAAQAAATEEAQALAIQA 1637

Qy     1114 GTDV-------TTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTT 1166
                    | || || |   |:      | :||       :  | :  :| :   ||
Db    1638 VLQAAQQAVMGTGEPMDTSEAAATV-----TQAEL----GHLSAEGQEGQATTIPIVLTQ 1688

Qy     1167 GTSASTSANSELVTSGS--------VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216
              |:     :|  : :          | | : || :  ::    | :    | |
Db    1689 QELAALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746

Qy     1217 LSQQVTSSSPSTNTFIA 1233
              |   | |   :|||:|
Db    1747 LPSTATESLAPSNTFVA 1763
```

```
RESULT 11
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910
;  GENERAL INFORMATION:
;    APPLICANT:  KLIS, FRANCISCUS M.
;    APPLICANT:  SCHREUDER, MAARTEN P.
;    APPLICANT:  TOSCHKA, HOLSER Y.
;    APPLICANT:  VERRIPS, CORNELIS T.
;    TITLE OF INVENTION:  PROCESS FOR IMMOBILIZING ENZYMES TO THE
;    TITLE OF INVENTION:  CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,525
;      FILING DATE:  04-JAN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92202080.5
;      FILING DATE:  08-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92203899.7
;      FILING DATE:  14-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP93/01763
;      FILING DATE:  07-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  213289/T7020(V)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-362-525-2


  Query Match             5.4%;  Score 350;  DB 3;  Length 650;
  Best Local Similarity   23.7%;  Pred. No. 1.1e-12;
  Matches  165;  Conservative  116;  Mismatches  292;  Indels  124;  Gaps   31;

Qy       5 FTLLFLYL---SIASAKTITGVFDSFNSLTWSN------AANYAFKGPGYPTWNAVLGWS 55
Db       2 FTFLKIILWLFSLALASAI-----NINDITFSNLEITPLTAN---KQPD-QGWTATFDFS 52

Qy      56 L-DGTSANPGDTFTLNMPCVFK---YTTSQTSVDLTADGVKYATCQFYSGE-----EFTT 106
Db      53 IADASSIREGDEFTLSMPHVYRIKLLNSSQTATISLADGTEAFKC-YVSQQAAYLYENTT 111

Qy     107 FSTLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKD 166
Db     112 F---TCTAQNDLSSYNTIDGSITFSLNFSDGGSSYEYELENAKFFKSGPMLVKLGNQMSD 168

Qy     167 ISIDVEFEKSTVDPSAY---LYASRVMPSLNKVTTLFVAPQCENGYTSG---TMGFSSSN 220

Db     169 V---VNF-----DPAAFTENVFHSGRSTGYGSFESYHLGMYCPNGYFLGGTEKIDYDSSN 220

Qy     221 GDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKY-QNVPAGYRPFIDAY 279
Db     221 NNVDLDCSSVQVYSSNDFNDWWFPQSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNAL 280

Qy     280 IS-ATDVNQYTLAYTNDYTC----AGSRLQSKPFTLR----WTGYKNSDAGSNGIVIVAT 330
Db     281 QSLPANVNTIDHALEFQYTCLDTIANTTYATQFSTTREFIVYQGRNLGTASAKSSFISTT 340

Qy     331 TRTVTD------STTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAP 384
Db     341 TTDLTSINTSAYSTGSISTV----ETGNRTTSEVISHVVTTSTKLSPTATTSLTIAQTSI 396

Qy     385 IGETATVIVDVPYHTTT-------TVTSEWTGTITTT-TTRTNPTDSIDTVVVQVPLPNP 436
Db     397 YSTDSNITVGTDIHTTSEVISDVETISRETASTVVAAPTSTTGWTGAMNTYIPQ------ 450

Qy     437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTV-TAPPG 495
Db     451 ------FTSSSFAT---------------INSTP--IISSSAVFETSDASIVNVHTENIT 487

Qy     496 GTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQS 555
Db     488 NTAAVPSEEP--TFVNATRNSLNSFCSSKQPSSPSSYTSS--------PLVSSLSVSKTL 537

Qy     556 YATTTTVTAPPGGTDSVIIREPP----NHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIR 611
Db     538 LSTSFTPSVP---TSNTYIKTENTGYFEHTALTTSSVGLNSFSETALSSQGTKIDTFLVS 594

Qy     612 E----PPNHTVTTTEYWSQSFATTTTVTGPPSGTDTV 644
Db     595 SLIAYPSSASGSQLSGIQQNFTSTSLMISTYEGKASI 631


RESULT 12
US-08-783-774-2
; Sequence 2, Application US/08783774
; Patent No. 6054130
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard
;    APPLICANT:  Jackman, Winthrop
;    TITLE OF INVENTION:  NON-SPLICING VARIANTS OF
;    TITLE OF INVENTION:  GP350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/783,774
;      FILING DATE:  15-JAN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7682-037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
```

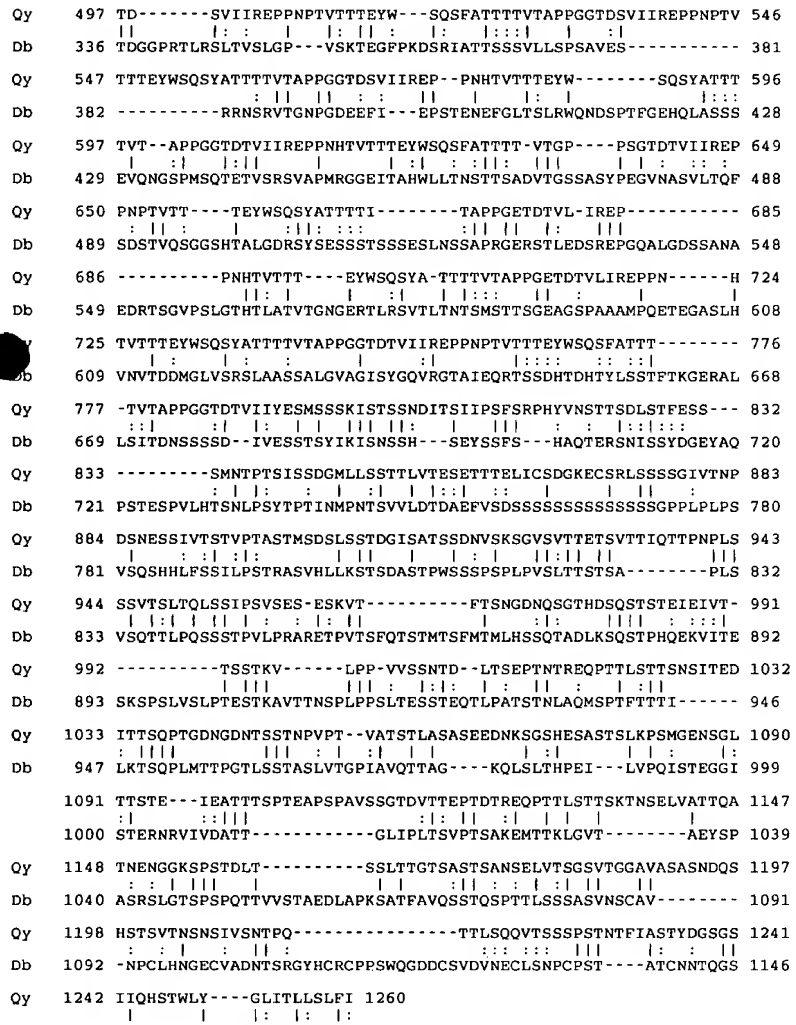
```
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-783-774-2


  Query Match             5.0%;  Score 323;  DB 3;  Length 907;
  Best Local Similarity   22.9%;  Pred. No. 6.1e-11;
  Matches  266;  Conservative  128;  Mismatches  390;  Indels  378;  Gaps   60;

Qy    152 TAGTN-TVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYT 210
          ||  | |: |: | |   :|::| : |    |                | ||:      |
Db     43 TADVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL 102

Qy    211 SGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPA 270
          : ||               : | :| |       || ||           : : :|:|
Db    103 ALTM-----------RSKKLPINVTTGEEQ---QVSLES-----------VDVYFQDV-- 135

Qy    271 GYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSK----PFTLRWTGYKNSDAGSNGIV 326
                                 |  : :|:     | |: :  : | :: :   |
Db    136 ----------------------FGTMWCHHAEMQNPVYLIPETVPYIKWDNCNSTNITAV 173

Qy    327 IVATTRTVT----------DSTTAVTTLPFNPSVDKTKTI---EILQPIP-------TTT 366
          : |   ||          ||  :| |      :|    :   || | :|          | :
Db    174 VRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEISQVLPGDNKFNITCS 233

Qy    367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
             |:|  :   || |:|:   || |     |  :  :|           :|    :||
Db    234 GYESHV-PSGGILTSTSPV---ATPIPGTGYAYSLRLTPR-------PVSRFLGNNSI-- 280

Qy    427 VVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486
          : |       |  | :|  ||              : |   | |          ||   |
Db    281 LYVFYSGNGPKASGGDYCIQS-------------NIVFSDEIPA---------SQDMPT 317

Qy    487 TTTVTAPPG--GTDSV--IIREPPN-PTVTTTEYWSQSFATTT------TVTA-PPGGTD 534
            ||      |  || :   | | || | :|:      | |      |:|:  | | :
Db    318 NTTDITYVGDNATYSVPMVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCE 377

Qy    535 SVIIREPPNPTVTTTEYWSQSYATTTT--VTAPPGGTDSVIIREPPNHTVTTTEYWSQSY 592
          ::                | ::|:  |  :|    ||       |    :| |
Db    378 NI----------------SGAFASNRTFDITVSGLGT------APKTLIITRT------- 408

Qy    593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPP-- 650
          ||  | |      |  ||  : |  | |:     :  ||    | || || |         |
Db    409 ATNATTT-----THKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAS 463

Qy    651 -NPTVTTTEYWSQSYATTTT----ITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTT 705
           |||:| :  | : | ||:   :| |  |     : |: |              ::|:
Db    464 TGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMT----------SSTSP 512

Qy    706 VTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAP-PGGTDTVIIREPPNPTVTT 764
          || |            || |           : |  || | |  |        | | |||
Db    513 VTTP----------TPNAT----------SPTPAVTTPTPNATS-------PTPAVTT 543

Qy    765 TEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTS 824
                |  |:|  |       ::  :| ::| : : ||  |:  :
Db    544 ---------PTPNATSPTLG-------KTSPTSAVTTPTPNATS--PTLGK--------- 576

Qy    825 DLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPD 884
            |  :|:: |||  :      :| ||                ||     :| :  || |
Db    577 ---TSPTSAVTTPTPNA------TSPTL-------------GK-----TSPTSAVTTPT 608

Qy    885 SNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSS 944
            |      :| ||    |   ::|         | :  |   |: | | | |  |   :|
Db    609 PN------ATGPTVGETSPQANAT---------NHTLGG---TSPTPVVTSQ--PKNATS 648

Qy    945 SVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSS 1004
          :||                              :| |:  |:|       |||  :  |   ||
Db    649 AVT--------------------------TGQHNITSSS-------TSSMSLRP---SS 671

Qy   1005 NTDLTSEPTNTREQPTTLS-TTSNSITED---ITTSQPTGDNGDNTSSTNPVPTVATSTL 1060
          |              | ||| :||:: |     :|:: |||  |:| :   |  :::|  :
Db    672 N------------PETLSPSTSDNSTSHMPLLTSAHPTG--GENITQVTPA-SISTHHV 715
```

```
Qy   1061 ASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATT-TSPTEAPSPAVSSGTDVTT 1119
               :::| |          :||     | :|  |   |:  |  | |  | ||   ||
Db    716 STSSPEP------RPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQAPSGQKTAV 769

Qy   1120 EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
                    ||  ||  | ||    ||       |||           |||  | ||
Db    770 --------PTVTSTGGKANS----TT------GGKH---------TTGHGARTSTEPTTD 802

Qy   1180 TSG-SVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDG 1238
               | |     :|:     |||          :|   | :|       |::    ::
Db    803 YGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNL--- 859

Qy   1239 SGSIIQHSTWLYGLITLLSLFI 1260
               |  ::| ::    ::||| | :
Db    860 SMLVLQWAS--LAVLTLLLLLV 879


RESULT  13
PCT-US95-04611A-19
; Sequence 19, Application PC/TUS9504611A
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  Non Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04611A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,291
;      FILING DATE:  April 18, 1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5163
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-04611A-19


  Query Match             5.0%;  Score 323;  DB 5;  Length 907;
  Best Local Similarity   22.9%;  Pred. No. 6.1e-11;
  Matches  266;  Conservative  128;  Mismatches  390;  Indels  378;  Gaps   60;

Qy    152 TAGTN-TVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYT 210
          ||  | |: |: | |   :|::| : |    |                | ||:      |
Db     43 TADVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL 102

Qy    211 SGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPA 270
          : ||               : | :| |       || ||           : : :|:|
Db    103 ALTM-----------RSKKLPINVTTGEEQ---QVSLES-----------VDVYFQDV-- 135
```

```
Qy    271 GYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSK----PFTLRWTGYKNSDAGSNGIV 326
Db    136 ----------------------FGTMWCHHAEMQNPVYLIPETVPYIKWDNCNSTNITAV 173

Qy    327 IVATTRTVT----------DSTTAVTTLPFNPSVDKTKTI---EILQPIP-------TTT 366
Db    174 VRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEISQVLPGDNKFNITCS 233

Qy    367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
Db    234 GYESHV-PSGGILTSTSPV---ATPIPGTGYAYSLRLTPR-------PVSRFLGNNSI-- 280

Qy    427 VVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486
Db    281 LYVFYSGNGPKASGGDYCIQS--------------NIVFSDEIPA---------SQDMPT 317

Qy    487 TTTVTAPPG--GTDSV--IIREPPN-PTVTTTEYWSQSFATTT------TVTA-PPGGTD 534
Db    318 NTTDITYVGDNATYSVPMVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCE 377

Qy    535 SVIIREPPNPTVTTTEYWSQSYATTTT--VTAPPGGTDSVIIREPPNHTVTTTEYWSQSY 592
Db    378 NI---------------SGAFASNRTFDITVSGLGT------APKTLIITRT------- 408

Qy    593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPP-- 650
Db    409 ATNATTT-----THKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAS 463

Qy    651 -NPTVTTTEYWSQSYATTTT----ITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTT 705
Db    464 TGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMT----------SSTSP 512

Qy    706 VTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAP-PGGTDTVIIREPPNPTVTT 764
Db    513 VTTP-----------TPNAT----------SPTPAVTTPTPNATS-------PTPAVTT 543

Qy    765 TEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTS 824
Db    544 ---------PTPNATSPTLG-------KTSPTSAVTTPTPNATS--PTLGK--------- 576

Qy    825 DLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPD 884
Db    577 ---TSPTSAVTTPTPNA------TSPTL-------------GK-----TSPTSAVTTPT 608

Qy    885 SNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSS 944
Db    609 PN------ATGPTVGETSPQANAT---------NHTLGG---TSPTPVVTSQ--PKNATS 648

Qy    945 SVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSS 1004
Db    649 AVT-------------------------TGQHNITSSS-------TSSMSLRP---SS 671

Qy   1005 NTDLTSEPTNTREQPTTLS-TTSNSITED---ITTSQPTGDNGDNTSSTNPVPTVATSTL 1060
Db    672 N-------------PETLSPSTSDNSTSHMPLLTSAHPTG--GENITQVTPA-SISTHHV 715

Qy   1061 ASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATT-TSPTEAPSPAVSSGTDVTT 1119
Db    716 STSSPEP------RPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQAPSGQKTAV 769

Qy   1120 EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
Db    770 --------PTVTSTGGKANS----TT------GGKH---------TTGHGARTSTEPTTD 802

Qy   1180 TSG-SVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDG 1238
Db    803 YGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNL--- 859

Qy   1239 SGSIIQHSTWLYGLITLLSLFI 1260
Db    860 SMLVLQWAS--LAVLTLLLLLV 879
```

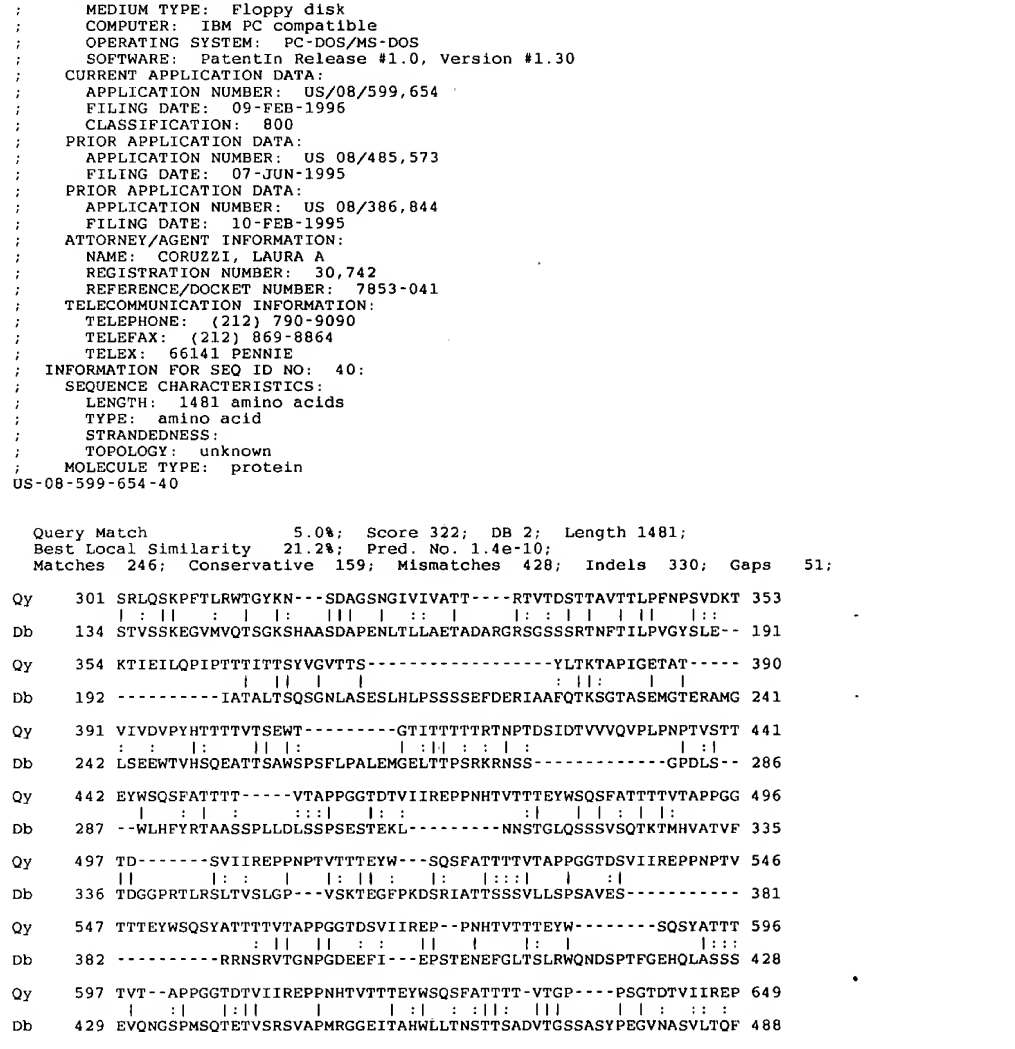
```
RESULT 14
US-08-616-844-40
; Sequence 40, Application US/08616844
; Patent No. 5849578
;  GENERAL INFORMATION:
;    APPLICANT:  FALB, DEAN A.
;    TITLE OF INVENTION:  COMPOSITION AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/616,844
;      FILING DATE:  15-MAR-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,654
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,573
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,844
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CORUZZI, LAURA A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-053
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-616-844-40
```

```
  Query Match             5.0%;  Score 322;  DB 2;  Length 1481;
  Best Local Similarity   21.2%;  Pred. No. 1.4e-10;
  Matches  246;  Conservative  159;  Mismatches  428;  Indels  330;  Gaps   51;

Qy    301 SRLQSKPFTLRWTGYKN---SDAGSNGIVIVATT----RTVTDSTTAVTTLPFNPSVDKT 353
Db    134 STVSSKEGVMVQTSGKSHAASDAPENLTLLAETADARGRSGSSSRTNFTILPVGYSLE-- 191

Qy    354 KTIEILQPIPTTTITTSYVGVTTS-----------------YLTKTAPIGETAT----- 390
Db    192 ----------IATALTSQSGNLASESLHLPSSSSEFDERIAAFQTKSGTASEMGTERAMG 241

Qy    391 VIVDVPYHTTTTVTSEWT---------GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT 441
Db    242 LSEEWTVHSQEATTSAWSPSFLPALEMGELTTPSRKRNSS------------GPDLS-- 286

Qy    442 EYWSQSFATTTT-----VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGG 496
Db    287 --WLHFYRTAASSPLLDLSSPSESTEKL---------NNSTGLQSSSVSQTKTMHVATVF 335
```



RESULT 15
US-08-599-654-40
; Sequence 40, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,654
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-599-654-40

Query Match 5.0%; Score 322; DB 2; Length 1481;
Best Local Similarity 21.2%; Pred. No. 1.4e-10;
Matches 246; Conservative 159; Mismatches 428; Indels 330; Gaps 51;



```
Qy    650 PNPTVTT----TEYWSQSYATTTTI--------TAPPGETDTVL-IREP---------- 685
Db    489 SDSTVQSGGSHTALGDRSYSESSSTSSSESLNSSAPRGERSTLEDSREPGQALGDSSANA 548

Qy    686 ----------PNHTVTTT----EYWSQSYA-TTTTVTAPPGETDTVLIREPPN------H 724
Db    549 EDRTSGVPSLGTHTLATVTGNGERTLRSVTLTNTSMSTTSGEAGSPAAAMPQETEGASLH 608

Qy    725 TVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTT-------- 776
Db    609 VNVTDDMGLVSRSLAASSALGVAGISYGQVRGTAIEQRTSSDHTDHTYLSSTFTKGERAL 668

Qy    777 -TVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESS--- 832
Db    669 LSITDNSSSSD--IVESSTSYIKISNSSH---SEYSSFS---HAQTERSNISSYDGEYAQ 720

Qy    833 ---------SMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNP 883
Db    721 PSTESPVLHTSNLPSYTPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSSGPPLPLPS 780

Qy    884 DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLS 943
Db    781 VSQSHHLFSSILPSTRASVHLLKSTSDASTPWSSSPSPLPVSLTTSTSA--------PLS 832

Qy    944 SSVTSLTQLSSIPSVSES-ESKVT----------FTSNGDNQSGTHDSQSTSTEIEIVT- 991
Db    833 VSQTTLPQSSSTPVLPRARETPVTSFQTSTMTSFMTMLHSSQTADLKSQSTPHQEKVITE 892

Qy    992 ----------TSSTKV------LPP-VVSSNTD--LTSEPTNTREQPTTLSTTSNSITED 1032
Db    893 SKSPSLVSLPTESTKAVTTNSPLPPSLTESSTEQTLPATSTNLAQMSPTFTTTI------ 946

Qy   1033 ITTSQPTGDNGDNTSSTNPVPT--VATSTLASASEEDNKSGSHESASTSLKPSMGENSGL 1090
Db    947 LKTSQPLMTTPGTLSSTASLVTGPIAVQTTAG----KQLSLTHPEI---LVPQISTEGGI 999

Qy   1091 TTSTE---IEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQA 1147
Db   1000 STERNRVIVDATT------------GLIPLTSVPTSAKEMTTKLGVT--------AEYSP 1039

Qy   1148 TNENGGKSPSTDLT----------SSLTTGTSASTSANSELVTSGSVTGGAVASASNDQS 1197
Db   1040 ASRSLGTSPSPQTTVVSTAEDLAPKSATFAVQSSTQSPTTLSSSASVNSCAV-------- 1091

Qy   1198 HSTSVTNSNSIVSNTPQ----------------TTLSQQVTSSSPSTNTFIASTYDGSGS 1241
Db   1092 -NPCLHNGECVADNTSRGYHCRCPPSWQGDDCSVDVNECLSNPCPST----ATCNNTQGS 1146

Qy   1242 IIQHSTWLY----GLITLLSLFI 1260
Db   1147 FICKCPVGYQLEKGICNLVRTFV 1169
```

Search completed: October 3, 2002, 15:32:32
Job time: 359 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:29:03 ; Search time 29.15 Seconds
(without alignments)
4153.433 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6495	100.0	1260	2	S60896	agglutinin-like pr
2	3194	49.2	1419	2	T30531	agglutinin-like ad
3	939	14.5	1367	2	S51959	hypothetical prote
4	939	14.5	1537	2	S53465	flocculation prote
5	711	10.9	1075	2	S48992	flocculation prote
6	675.5	10.4	1367	1	S48478	glucan 1,4-alpha-g
7	649.5	10.0	1275	2	T33369	hypothetical prote
8	646.5	10.0	2232	2	T34434	hypothetical prote
9	637.5	9.8	2271	2	F90073	hypothetical prote
10	632.5	9.7	3570	2	T45025	mucin MUC5B, trach
11	599	9.2	3020	2	A43932	mucin 2 precursor,
12	578	8.9	1283	2	T39174	hypothetical Serin
13	577	8.9	1161	2	S57180	probable membrane
14	553.5	8.5	1459	2	T32271	hypothetical prote
15	547.5	8.4	1609	2	S25345	probable membrane
16	526	8.1	1229	2	T25697	hypothetical prote
17	526	8.1	1630	2	A53577	ascites sialoglyco
18	515	7.9	4776	2	E95206	cell wall surface
19	511.5	7.9	3507	2	T34513	hypothetical prote
20	503.5	7.8	1777	2	T34369	hypothetical prote
21	501	7.7	1169	2	S38181	flocculation prote
22	492.5	7.6	1104	2	S59310	probable membrane
23	487	7.5	1832	2	T31113	mucin-like glycopr
24	482.5	7.4	1829	2	T24583	hypothetical prote
25	464.5	7.2	670	2	F36791	hypothetical prote
26	460	7.1	770	2	T22808	hypothetical prote
27	451.5	7.0	786	2	T16509	hypothetical prote
28	445	6.9	1802	2	S69703	HKR1 protein precu
29	443.5	6.8	1032	2	T34433	hypothetical prote
30	443.5	6.8	1251	2	T21389	hypothetical prote
31	435.5	6.7	2468	2	A83412	hypothetical prote
32	430.5	6.6	866	2	T45462	membrane glycoprot
33	428	6.6	867	2	T45463	membrane glycoprot
34	424	6.5	528	2	I47141	gastric mucin (clo
35	423.5	6.5	520	2	S62521	hypothetical prote
36	419	6.5	973	2	T40778	hypothetical 129.5
37	418.5	6.4	725	2	A41258	a-agglutinin core
38	415.5	6.4	1306	2	S25370	MSB2 protein - yea
39	412.5	6.4	3029	2	S76109	hypothetical prote
40	412	6.3	851	2	T22696	hypothetical prote
41	409.5	6.3	1041	2	S55862	probable membrane
42	405	6.2	833	2	E90577	lipoprotein vsaC [
43	404.5	6.2	2035	2	A40718	host cell factor C
44	401.5	6.2	648	2	PC4395	mucin 3 - human (f
45	400.5	6.2	948	2	T11678	hypothetical prote

ALIGNMENTS

RESULT 1
S60896
agglutinin-like protein - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-2000
C;Accession: S60896
R;Hoyer, L.L.; Scherer, S.; Shatzman, A.R.; Livi, G.P.
Mol. Microbiol. 15, 39-54, 1995
A;Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual
A;Reference number: S60896; MUID:95272392
A;Accession: S60896
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1260 <HOY>
A;Cross-references: EMBL:L25902; NID:g704426; PIDN:AAC41649.1; PID:g704427
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 100.0%; Score 6495; DB 2; Length 1260;
Best Local Similarity 100.0%; Pred. No. 1.7e-273;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Qy    481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db    481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy    541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db    541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Qy    601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db    601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Qy    661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db    661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720

Qy    721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db    721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy    781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
Db    781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840

Qy    841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
Db    841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900

Qy    901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960
Db    901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960

Qy    961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020
Db    961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020

Qy   1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
Db   1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080

Qy   1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140
Db   1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140

Qy   1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200
Db   1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200

Qy   1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db   1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
```

RESULT 2
T30531
agglutinin-like adhesin - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C;Accession: T30531
R;Gaur, N.K.; Klotz, S.A.
Infect. Immun. 65, 5289-5294, 1997
A;Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, tha
A;Reference number: Z20847; MUID:98053977
A;Accession: T30531
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1419 <GAU>
A;Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDN:AAB88883.1
C;Genetics:
A;Gene: ALA1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

```
  Query Match            49.2%;  Score 3194;  DB 2;  Length 1419;
  Best Local Similarity  49.5%;  Pred. No. 9.8e-131;
  Matches  708;  Conservative  182;  Mismatches  355;  Indels  186;  Gaps   29;

Qy      1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db      1 MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTS 60

Qy     61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db     61 ANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVNNNLRS 120

Qy    121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db    121 SIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGSKKLSIAVNFEKSTVDQ 180

Qy    181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db    181 SGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVND 240

Qy    241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db    241 WNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVD 300

Qy    301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db    301 DYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy    361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db    361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy    421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db    421 TDSIDTVVVQVPSPNPTTTTTQFWSESFTSTTTITNSLKGTDSVIVREPHNPTVTTTEFW 480

Qy    481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db    481 SESFATTETITSKPEGTDSVIVREPHNPTVTTTEFWSESYATTETITNGPEGTDSVIVRE 540

Qy    541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db    541 PHNPTVTTTKFWSESYATTETITNKPEGTDSVIVKEPYNPTVTTTEFWSESYATTETITN 600

Qy    601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTT--- 657
Db    601 GPEGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEESSSTTAIE 660

Qy    658 ----------EYWSQSYATTTTITAPPGETDTVLIREPPNHTVT---------------- 691
Db    661 SSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYDSD 720

Qy    692 ----------------------TTEYWSQS----YATTTTVTAPPGETDTVLIREPPNHT 725
Db    721 SSSSIESSTLSSSDRCSSSISDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNSI 780

Qy    726 VTTTEYWSQSYATTTTVTAPPGGTDTVIIREP------------PNPTVTTTEYW----- 768
Db    781 STSQELSSSSSEEESSTF-----ATDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIA 835

Qy    769 ------SQSFATTT--------TVTAPPGGT-DTVIIYESMSSSKISTSSNDI------- 806
Db    836 GESDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFS 895

Qy    807 TSIIPSFS----------RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTES 856
Db    896 TLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESAN--DDTYTLSES 953

Qy    857 ETTTELICSDGKE---CSRLSSSSGIVTN-PDSNESSIVTSTVPTASTMSDSLSSTDG-I 911
Db    954 TDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSAEQ 1013
```

```
Qy    912 SATSSDNV--SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTS 969
Db   1014 SITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETS--TLSSDDSTS-SDTSISSTT 1070

Qy    970 NGD----NQSGTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSEPTNTR------ 1016
Db   1071 NSDTGNINAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELITTEL 1130

Qy   1017 -----EQPTTLSTTSNSITE------DITTSQPTGDNGDNTSSTNPVPTVAT----STLA 1061
Db   1131 TTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTEA 1190

Qy   1062 SA-----SEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTD 1116
Db   1191 STATLGENSFSNVASTPSNIATSLSRSTSSSSNHATESSGTVKSEASAEAIPSPPTS--TD 1248

Qy   1117 VTTEPTDTREQPTTLSTTSKTN-----SELVATTQATN---ENGGKSPSTDLTSSLTTGT 1168
Db   1249 NRLSYSTEEAKGITYANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQ 1308

Qy   1169 SASTSANSELV----TSGSVTGGAVASASNDQSHSTSV----------------TNSNSI 1208
Db   1309 PSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSA 1368

Qy   1209 VSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
Db   1369 ASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKHPSWLLKFISVALFF 1417
```

RESULT 3
S51959
hypothetical protein YAL063c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 22-Oct-1999
C;Accession: S51959
R;Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A;Reference number: S51956
A;Accession: S51959
A;Molecule type: DNA
A;Residues: 1-1367 <BUS>
A;Cross-references: EMBL:U12980; GSPDB:GN00001; MIPS:YAL063c
C;Genetics:
A;Gene: MIPS:YAL063c
A;Map position: 1L

Query Match 14.5%; Score 939; DB 2; Length 1367;
Best Local Similarity 28.1%; Pred. No. 2.9e-33;
Matches 410; Conservative 214; Mismatches 469; Indels 364; Gaps 84;

```
Qy     28 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQT----S 83
Db     50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCV---SSSGTFPCPQ 98

Qy     84 VDLTAD-GVK-YATCQ------FYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIA-- 133
Db     99 EDLYGNWGCKGIGACSNNPIIAYWSTDLFGFYTTPT--------------NVTLEMTGY 143

Qy    134 FNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSL 193
Db    144 FLPPQTGSYT----------------------------FKFATVDDSAIL--------- 165

Qy    194 NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC----------SNIHIGITKGLNDWN- 242
Db    166 ------------------------SVGGSIAFECCAQEQPPITSTNFTI---NGIKPWNG 198

Qy    243 -YPVSSESFSYTKTCTSNGIQIKYQNVPA-GYRPF------------IDAYISATDVN- 286
Db    199 SPPDNITGTVYMYAGFYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEGYVYTFDNNL 258

Qy    287 ---QYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDA-------GSNGI-----VIVATT 331
Db    259 SQPNCTIPDPSNYTVSTTITTTEP----WTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy    332 RTVTDSTTAVTTLPFNPSVDKTKT----------------IEILQP-IPTTTITT---- 369
Db    315 PT-TASTIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETIIVIRTPTTATTAITTTEPW 373

Qy    370 --SYVGVTTSYLTKTA----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTT 415
Db    374 NSTFTSTSTELTTVTGTNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTV 432

Qy    416 TRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQSFATT----TTVTAPPG-GTD--TVI 465
Db    433 TGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGLPTDETIIV 491

Qy    466 IREPPNHT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNPT--VTTTE 514
Db    492 IRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPTTATTAMTTTQ 551

Qy    515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNPT--VTTTEYWSQSYATT----TT 561
Db    552 PWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTEMTT 611

Qy    562 VTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQSYATT----TTVTAPPG-GTD--TVI 609
Db    612 VTGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIV 671

Qy    610 IREPPNHT--VTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPT--VTTTE 658
Db    672 IRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPTTATTAMTTTQ 731

Qy    659 YWSQSYATT----TTITAP---PGETDTVLIREPPNHTV--TTTEYWSQSYATT----TT 705
Db    732 PWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTT 791

Qy    706 VTAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPG-GTD--TVI 753
Db    792 VTGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIV 851

Qy    754 IREPPNPTV--TTTEYWSQSFATT----TTVTAPPG--GTDTVIIYESMS---SSKISTS 802
Db    852 IRTPTSEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIIVKTPTTAISSSLSSS 911

Qy    803 SNDITSIIPS---FSRPHYVNSTTSDLSTFESSSMNTPTSISSDGM---LLSSTTLVTES 856
Db    912 SGQITSFITSARPIITPFYPSNGTSVISSSVISSSDTSSLVISSSVTSSLVTSSPVISSS 971

Qy    857 ETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSS 916
Db    972 FISSPVISSTTTSASILSESS------------KSSVIPTSSSTSGSSESETGSASSAS 1018

Qy    917 DNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSG 976
Db   1019 ---SSSSISSESPKSTYSSSSLP-PVTSATTSQEITSSLPPVTTTKT------------ 1061

Qy    977 THDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPT---NTREQPTTLSTTSNSITEDI 1033
Db   1062 -----SEQTTLVTVTSCESHVCTESISSAIVSTATVTVSGATTEYTTWCPISTTEITKQT 1116

Qy   1034 T-TSQPTGDNGDNTSSTNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGL 1090
Db   1117 TETTKQTKGTTEQTTETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEY 1173

Qy   1091 TTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVAT-----T 1145
Db   1174 TTWCPISTTESKQQTTLVTVTSCGSGVCSETT----SPAIVSTATATVNDVVTVYSTWRP 1229

Qy   1146 QATNENGGKSPSTDLTSSLTTGTSAST------SANSELVTSGSVT-----GGAVASASN 1194
Db   1230 QTTNEQSVSSKMNSATSETTTNTGAAETTTSTGAAETKTVVTSSISRFNHAETQTASATD 1289

Qy   1195 DQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS------SPSTNTFIASTYDGSGSII 1243
```

```
              ||:||     | :   :::: :|:||| |:        || :   ||| || : :
Db   1290 VIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGSSTASLEISTYAGSANSL 1349

Qy   1244 QHSTWLYGLITLLSLFI 1260
             : |   |  | | |
Db   1350 LAGSGLSVFIASLLLAI 1366


RESULT   4
S53465
flocculation protein FLO1 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YAR050w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C;Accession: S53465; S43543; S57851; S31230; S57349
R;Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac
submitted to the EMBL Data Library, February 1994
A;Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 5
A;Reference number: S53458
A;Accession: S53465
A;Molecule type: DNA
A;Residues: 1-1537 <BUS>
A;Cross-references: EMBL:L28920; NID:g1616966; PIDN:AAC09499.1; PID:g694125; MIPS:YAR050
R;Watari, J.; Takata, Y.; Ogawa, M.; Sahara, H.; Koshino, S.; Onnela, M.L.; Airaksinen,
Yeast 10, 211-225, 1994
A;Title: Molecular cloning and analysis of the yeast flocculation gene FLO1.
A;Reference number: S43543; MUID:94262325
A;Accession: S43543
A;Molecule type: DNA
A;Residues: 1-428,'M',430-473,'M',475-518,'M',520-549,'T',551-608,'L',610-636,'M',638-69
A;Cross-references: EMBL:X78160
R;Onnela, M.
submitted to the EMBL Data Library, September 1994
A;Reference number: S57851
A;Accession: S57851
A;Molecule type: DNA
A;Residues: 1-428,'M',430-463,'D',465-473,'M',475-518,'M',520-549,'T',551-608,'L',610-63
A;Cross-references: EMBL:X78160; NID:g535933; PIDN:CAA55024.1; PID:g535934
R;Teunissen, A.W.R.H.; Holub, E.; van der Hucht, J.; van den Berg, J.A.; Steensma, H.Y.
Yeast 9, 423-427, 1993
A;Title: Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerevisi
A;Reference number: S31230; MUID:93289821
A;Accession: S31230
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-296,927-1516,'TAYWPVVV' <TEU>
R;Bidard, F.; Bony, M.; Blondin, B.; Dequin, S.; Barre, P.
Yeast 11, 809-822, 1995
A;Title: The Saccharomyces cerevisiae FLO1 flocculation gene encodes for a cell surface
A;Reference number: S57349; MUID:96090130
A;Accession: S57349
A;Molecule type: DNA
A;Residues: 1243-1274;1308-1339;1359-1390 <BID>
C;Genetics:
A;Gene: SGD:FLO1
A;Cross-references: SGD:S0000084; MIPS:YAR050w
A;Map position: 1R
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1537/Product: flocculation protein FLO1 #status predicted <MAT>
F;953-997/Domain: repeat A2 <RA2>
F;998-1042/Domain: repeat A3 <RA3>
F;1043-1081/Domain: repeat A4 <RA4>
F;1226-1276/Domain: repeat B1 <RB1>
F;1277-1284/Domain: repeat B2 (partial) #status atypical <RB2>
F;1291-1341/Domain: repeat B3 <RB3>
F;1342-1392/Domain: repeat B4 <RB4>
F;1408-1416/Domain: repeat C1 <RC1>
F;1417-1425/Domain: repeat C2 <RC2>
F;1426-1434/Domain: repeat C3 <RC3>
F;135,187,262,1114/Binding site: carbohydrate (Asn) (covalent) #status predicted
```

```
  Query Match             14.5%;  Score 939;  DB 2;  Length 1537;
  Best Local Similarity   25.8%;  Pred. No. 3.3e-33;
  Matches  414;  Conservative  246;  Mismatches  520;  Indels  426;  Gaps   84;

Qy       2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
           |  |||| |  |:||  |           :|:  :|      :| |:|||  |:
Db      10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy      41 ----------------------------KGPGYPTW----------NAVLGWSLD--G 58
                                         :  |  |          :  || |  |
Db      69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy      59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
             | :  || |             ||: |                          || :
Db     129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy      86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
           | ||:|             |   |:|  :             : |:  ::||    ||||
Db     188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy     132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
                ||  | | |           :|:|       |:     || ||| |   |
Db     239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy     191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
           :    | | :  |    :|| | :   : |          |     ||    :| |
Db     278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy     251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
             |   :||::     :       |    :||      |  : : :|   : |
Db     338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy     311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT----------- 355
                 :  |:||:     :   || | :|||:||  |:| :   | |
Db     385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGL 438

Qy     356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
                 | |  |  ||  |         ::   :|   | |     |  ||   :::  |
Db     439 PTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGLPTDET-IIVIRTPTT 497

Qy     399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPTVSTT-----EY 443
             ||  | | |  | |       || | ||  |||  :|:::|   :  || :||     :
Db     498 ATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTD--ETIIV---IRTPTTATTAMTTPQP 552

Qy     444 WSQSFATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTV 490
           |: :| :|    ||||   |  ||   ::|| |   |  :|||| |: :| :|    |||
Db     553 WNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTEMTTV 612

Qy     491 TAPPG-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VII 538
           |   |  ||   ::|| |   |  :|||: |: :| :|    ||||   |  ||   ::|
Db     613 TGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVI 672

Qy     539 REPPNPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEY 587
           | |   |  :|||: |: :: :|    ||||  |  ||   ::|| |   |  :|||:
Db     673 RTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTTGLPTDETIIVIRTPTTATTAMTTTQP 732

Qy     588 WSQSYATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTV 634
           |: :: :|    ||||   |  ||   ::|| | :  :  |||| |: :| :|    |||
Db     733 WNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTV 792

Qy     635 T---GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLI 682
           |   | |:    ::|| | :    |||| |: :: :|    ||||     |  :   ::|
Db     793 TGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVI 852

Qy     683 REPPNHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEY 731
           | |  :  :  |||| |: :: :|    ||:|   |: ||   ::|| | :  :  ||||
Db     853 RTPTSEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEP 912

Qy     732 WSQSYATTTTVTAPPGGTD-------TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPP 782
           |: :: :|:|      ||:        ::||.| :  :  |||| |: :| :|:|
Db     913 WTGTFTSTSTEMTHVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 972

Qy     783 GGT------DTVIIYESMSSSKI-----------STSSNDITSIIPSFSRP-------- 816
```

```
Db    973 TGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 1032

Qy    817 ------HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICS 865
Db   1033 RTPTSEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSS 1092

Qy    866 DGKECSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSG 923
Db   1093 SGQITSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTS 1151

Qy    924 VSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST 983
Db   1152 TSIFSESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS- 1204

Qy    984 STEIEIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITE 1031
Db   1205 ---LPLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISP 1252

Qy   1032 DITTSQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSG 1089
Db   1253 AIVSTATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCS 1312

Qy   1090 LTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELV 1142
Db   1313 KTASPAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAI 1370

Qy   1143 ATT----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSV 1184
Db   1371 VSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITN 1430

Qy   1185 TGGA-----------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQ 1220
Db   1431 TGAAETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQ 1490

Qy   1221 VTSSSP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db   1491 PRSTPASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536
```

RESULT 5
S48992
flocculation protein homolog YHR211w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 29-Oct-1999
C;Accession: S48992
R;Macri, C.
Submitted to the EMBL Data Library, February 1994
;Description: The sequence of S. cerevisiae cosmid 9177.
A;Reference number: S46671
A;Accession: S48992
A;Molecule type: DNA
A;Residues: 1-1075 <MAC>
A;Cross-references: EMBL:U00029; NID:g551322; PID:g458919; GSPDB:GN00008; MIPS:YHR211w
C;Genetics:
A;Gene: MIPS:YHR211w
A;Map position: 8R

Query Match 10.9%; Score 711; DB 2; Length 1075;
Best Local Similarity 25.9%; Pred. No. 1.5e-23;
Matches 324; Conservative 192; Mismatches 393; Indels 342; Gaps 66;

```
Qy    153 AGTNTVTFNDGDKDISIDVEF------EKSTVDPSAYL---YASRV-------------- 189
Db     23 SGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKTKLGSVGGQTDISID 82

Qy    190 --MPSLNKVTTLFVAPQCENGYTS-GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
Db     83 YNIPCVSSSGT-FPCPQ-EDSYGNWGCKGMGA--------CSN-----SQGIAYW----S 123

Qy    247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR---L 303
```

```
Db    124 TDLFGFYTTPTNVTLEMTGYFLPPQTGSYTFSFATVDDSAILSVGGSIAFECCAQEQPPI 183

Qy    304 QSKPFTLRWTGYKNSDAGSNGIVIVATT----------RTVTDSTTAVTTLPFNPSVDKT 353
Db    184 TSTNFTI--NGIKPWD-GSLPDNITGTVYMYAGYYYPLKVVYSNAVSWGTLPI--SVE-- 236

Qy    354 KTIEILQPIPT-TTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT---TTVTSEWTG 409
Db    237 --------LPDGTTVSDNFEGYVYSF---DDDLSQSNCTIPDPSIHTTSTITTTTEPWTG 285

Qy    410 TIT------TTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDT 463
Db    286 TFTSTSTEMTTITDTNGQLTDETVIV---IRTPT------------TASTIT-------- 322

Qy    464 VIIREPPNHTVTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTE 514
Db    323 -----------TTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTE 371

Qy    515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTEYWSQSYATTTTVTAP 565
Db    372 PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEVTT 431

Qy    566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNH--TVTTTEY 623
Db    432 ITGTNG----QPTDETV------------------------IVIRTPTSEGLITTTTEP 462

Qy    624 WSQSFATT----TTVT---GPPSGTDTVIIREPPNPTV--TTTEYWSQSYATT----TTI 670
Db    463 WTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 522

Qy    671 TAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPGE-TD--TVLI 718
Db    523 TGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 582

Qy    719 REPPNH--TVTTTEYWSQSYATTTTVTAPPGGT------DTVIIREPPNPTVTTTEYWSQ 770
Db    583 RTPTSEGLITRTTEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPT-TAISSSLSSS 641

Qy    771 SFATTTTVTA---------PPGGTDTV---IIYESMSSSKISTSSNDITSIIPSFSRPHY 818
Db    642 SGQITSSITSSRPIITPFYPSNGTSVISSSVISSSVTSSLVTSSSFISSSVISS------ 695

Qy    819 VNSTTSDLSTFESSSMNT--PTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
Db    696 --STTTSTSIFSESSTSSVIPTSSSTSGSSESKTSSASSSSSSSSISSESPKSPTNSSSS 753

Qy    877 SGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS------SDNVSKSGVSVTTET 930
Db    754 LPPVTSATTGQET--ASSLPPATTTKTS-EQTTLVTVTSCESHVCTESISSAIVSTATVT 810

Qy    931 --SVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE 988
Db    811 VSGVTTEYTTWCPISTTETT----------------------KQTKGTTEQTKGTTEQT 847

Qy    989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048
Db    848 TETTKQTTVV-TISSCESDICSKTAS----PAIVSTSTATI-----------NGVTTEY 890

Qy   1049 TNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP- 1107
Db    891 TTWCPISTTES-----------------------KQQTTLVTVTSCESGVCSETTSPA 925

Qy   1108 --SPAVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL 1160
Db    926 IVSTATATVNDVVTVYPTWRPQTTNEQ----SVSSKMNS---ATSETTTNTGAAETKTAV 978

Qy   1161 TSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSV-----TNSNSIVSNTPQT 1215
Db    979 TSSLSRFNHAETQ--------------TASATDVIGHSSSVVSVSETGNTMSLTSSGLS 1023

Qy   1216 TLSQQVTSS------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db   1024 TMSQQPRSTPASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1074
```

```
RESULT   6
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009; MIP
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281; MUID:89031230
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123; MUID:96323237
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>


  Query Match            10.4%;  Score 675.5;  DB 1;  Length 1367;
  Best Local Similarity  26.2%;  Pred. No. 7.1e-22;
  Matches  378;  Conservative  177;  Mismatches  535;  Indels  353;  Gaps   65;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           | : | | :| ||:            |||              |:||     | | :|||
Db       1 MQRPFLLAYLVLSLL-----------FNSAL-----------GFPTALVPRG-SSEGTS 36

Qy      61 AN------PGDTFTLNM--PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTC 112
             |      |   | :|    : :||   |||    |   |       |:|
Db      37 CNSIVNGCPNLDFNWHMDQQNIMQYTLDVTSVSWVQDNT-YQITIHVKGKE--------- 86

Qy     113 TVNDALKS--SIKAFGTVTLPIAFNVGGTGSSTDLEDSKC-FTAGTNTVTFN-DGDKDI- 167
             |  ||   |:| |       : |   :| | |:   |||     ||     :|:
Db      87 --NIDLKYLWSLKIIGVTGPKGTVQLYGYNENTYLIDNPTDFTA-----TFEVYATQDVN 139

Qy     168 -------SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 220
                  :  ::||        ||  |              | |  : :  ||  |  |

Db     140 SCQVWMPNFQIQFE--------YLQGS--------------AAQYASSWQWGTTSFDLST 177

Qy     221 GDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI 280
           |    |:|      :| :  ::|     | :   | :|
Db     178 G-----CNNYD---NQGHSQTDFP----GFYWNIDCDNN-------------------C 205

Qy     281 SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA 340
             |  : |: :   |   | :|   |       :| : |:      |: |   ||::
Db     206 GGTKSSTTTSSTSESSTTTSSTSESSTTT-------SSTSESS----TTTSSTSESSTSS 254

Qy     341 VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT 400
            || |  |:     |    :  ||   |||          ||  |          |:| |
Db     255 STTAPATPT-----TTSCTKEKPTPPTTTS--------CTKEKP---------TPPHHDT 292

Qy     401 TTVTSEWTGTITTTTTRTN---PTDSIDTV-VVQVPLPNPTVSTTEYWSQ--SFATTTTV 454
           |  | : | |  | | :|    || |  |       |:| |: |||| |    : :|| :
Db     293 TPCTKKKTTTSKTCTKKTTTPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESS 352

Qy     455 TAP---PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVT 511
           :||   |  : |     |   : | :     : :|| : :||           | |: :
Db     353 SAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAP----------VPTPSSS 401

Qy     512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTT-VTAPPGGTD 570
           |||  |    ::|| ::    | |         | :|||  |    ::||  :: |  |
Db     402 TTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTP 461

Qy     571 SVIIREPPNHTVT--TTEYWS-----QSYATTTTVTAPPGGTDTVIIREP-PNHTVTTTE 622
           |    |  :  || |||  |       | :|| : :||   : |     | |  : :|||
Db     462 SSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE 521

Qy     623 YWS-----QSFATTTTVTGPPSGTDTVIIREP-PNPTVTTTEYWSQSYATTTTITAPPGE 676
           |      | :|| : : | : : |      | | |: :|||      :::| :|:   |
Db     522 SSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE------SSSTPVTSSTTE 575

Qy     677 TDTVLIREPPNHTVTTTEYWSQSYAT----TTTVTAPPGETDTVLIREPPNHTVT--TTE 730
           : :  :   |  : :|||  |    |    ||  :: |  | :    |  :  ||  |||
Db     576 SSSAPV---PTPSSSTTESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTE 632

Qy     731 YWS-----QSYATTTTVTAP---PGGTDTVIIREP-PNPTVTTTEYWSQSFATTTTVTAP 781
           |      | :|| : :||   |  : |     | | |: :|||  | |   |:: |
Db     633 SSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE--SSSAPVTSSTTES 690

Qy     782 PGGTDTVIIYESMSS-----SKISTSSNDITSIIPSFSR--------PHYVNSTTSD--- 825
                |    || |:     |  :| |:      || |          |   :|||
Db     691 SSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSA 750

Qy     826 ---LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
              || ||||   ||  ||     ||  : | | :|||   :     |  :: | :
Db     751 PVTSSTTESSSAPVPTP-SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPV 809

Qy     883 PDSNESSIVTSTVPTASTMSDSL-SSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNP 941
           |  : || :||: |:::  | |  ||:  :   ||      |   |:: |: :::   | |
Db     810 PTPSSSSNITSSAPSSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAPVPTP 869

Qy     942 LSSSVTSLTQLSSIPSVSESES-----KVTFTSNGDNQSGTHDSQSTSTEIEIV------ 990
           |||  : :  ||||  | :||      || :|:    | |  | |::||  ||
Db     870 SSSSNITSSAPSSIPFSSTTESFSTGTTVTPSSSKYPGSQTETSVSSTTETTIVPTKTTT 929

Qy     991 ---TTSSTKVLPPVVSSNTDLTSEPTN-------TREQPTTLST--TSNSITEDITTSQP 1038
             | |:| :   | |: |:   | |:      |   ||| :|  |::| |   ||
Db     930 SVTTPSTTTITTTVCSTGTNSAGETTSGCSPKTVTTTVPTTTTTSVTTSSTTTITTTVCS 989

Qy    1039 TGDN--GDNTSSTNPVPTVATSTLASASEEDNKSGSHESA-----------------ST 1078
           || |  |: ||  :|  |: |:   | |  :  | |  ::                 ||
Db     990 TGTNSAGETTSGCSP-KTITTTVPCSTSPSETASESTTTSPTTPVTTVVSTTVVTTEYST 1048

Qy    1079 SLKPSMGENSGLTTS--TEIEATTTSPTEAPSPAV----------------SSGTD---- 1116
           | ||  ||   :||:  |:   ||   | ||:|:|                :|||:
Db    1049 STKPG-GE---ITTTFVTKNIPTTYLTTIAPTPSVTTVTNFTPTTITTTVCSTGTNSAGE 1104

Qy    1117 ----------VTTEPTDT-----REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLT 1161
                     || |  |       : ||| ||: | : :| |  :|  |     :|  |
Db    1105 TTSGCSPKTVTTTVPCSTGTGEYTTEATTLVTTAVTTT-VVTTESSTGTNSAGKTTTGYT 1163
```

```
Qy   1162 SSLTTGTSASTSANSELVTSG------SVTGGAVASASNDQSHSTSVTNSNSIVSN---- 1211
Db   1164 TKSVPTTYVTTLAPSAPVTPATNAVPTTITTTECSAATNAAGETTSVCSAKTIVSSASAG 1223

Qy   1212 --------TPQTT---LSQQVTSSSPSTN---------------TFIASTYDGSGSIIQ 1244
Db   1224 ENTAPSATTPVTTAIPTTVITTESSVGTNSAGETTTGYTTKSIPTTYITTLIPGSNGAKN 1283

Qy   1245 HST 1247
Db   1284 YET 1286
```

RESULT 7
T33369
hypothetical protein H02F09.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C;Accession: T33369
R;Geisel, C.; Harmon, G.
Submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid H02F09.
A;Reference number: Z21330
A;Accession: T33369
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1275 <GEI>
A;Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
A;Experimental source: strain Bristol N2; clone H02F09
C;Genetics:
A;Gene: CESP:H02F09.3
A;Map position: X
A;Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 10.0%; Score 649.5; DB 2; Length 1275;
Best Local Similarity 26.8%; Pred. No. 8.6e-21;
Matches 326; Conservative 151; Mismatches 449; Indels 289; Gaps 53;

```
Qy    213 TMGFSSSNG-----------DVAIDCS------NIHIGITKGLNDWNYPVSSESFSYTKT 255
Db     49 TVSLCNSNSEPQISSLRFKRDANTDCSFNANKNQLRNFLTSSAEDANFILTTYSGTSTKT 108

Qy    256 --------------CTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db    109 EPLTKQLALDQLEAIKADGSSAVKQSSAIESYDP--SAY-SNTDLVFFTPCQTN------ 159

Qy    301 SRLQSKPFTLRWTG----YKNSDAGSNGI-----VIVATTRTVTDSTT----AVTTLPFN 347
Db    160 -----------YDGIEDDIKNVQTAINKVITKTFVIVSLSLNSTDMNSRYGEAAHNIPTT 208

Qy    348 PSVDKTKTIEILQPIPTT---TITTSYVGVTTSYLTKTAP-----IGETATVIVDVPYHT 399
Db    209 PTEDISNKINNILNIGTTQTPPVTTSTMATTTANVTSAAPNTTVTISTSPTTVVTVP--- 265

Qy    400 TTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPG 459
Db    266 STAQTSSTTTVTVPTTTVTGP-----TTVVTVPTTVVTIPSTVVTSPITTPSTVVTVP-- 318

Qy    460 GTDTVIIREPPNHTVTTTEYWSQSFATT---TTVTAPPGGTDSVIIREP-----PNPTVT 511
Db    319 --STVV-------TVPSTAVTKPSTVVTAPSTVVTVP-----STVVTKPNTVVTSSPTVA 364

Qy    512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPN-PTVTTTEYWSQSYATTTTVTAPPGGTD 570
Db    365 TT--------PTTVVTTP-----STVVTVPSTVVTVPTTVVTNPS----TVVTAP----- 402

Qy    571 SVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTE-------- 622
Db    403 STVV------TVPTTVMTSRSTVITTPTT---GGSSPSTAGTSLASTAVTTETSIGSSST 453

Qy    623 -------YWSQSFATTTTVTGPPSGTDTVIIREPPNPTV---------------TTTEY 659
Db    454 PLPSQSTSLSMSSLSTYTPSSSTAGATSPATQQSTKPTIGTSMSSGPTTVAPGASTESTV 513

Qy    660 WSQSYATTTTITAPPGETDTVLIREPPNHTVTT-TEYWSQSYATTTTVTAPPGETDTVLI 718
Db    514 LQSSTPSGTTVTLPSGSSTATAGTSPQASTVTTVTDISTVSGSTVTSQTA-----ESSLS 568

Qy    719 REPPN------HTVTTTEYWSQSY------------ATTTTVTAPPGGTDTVIIREPPN 759
Db    569 TESPTSAGSSISTVSTVSSQPSTYIPVSSASSIYSTLSGSTGSTASPGTTESSGSSTSGP 628

Qy    760 PTVTTTEYWSQSFATTTTVTAP------------PGGTD-TVIIYESMSSSKISTSSNDI 806
Db    629 STISGS---SASTVTGSTVTEASTISGSTESSTIPGSTESTVSEASTVSGSSVSTVSGST 685

Qy    807 TSIIPSFSR-PHYVNSTTSDLSTFESS--SMNTPTSI----------------SSDGML 846
Db    686 ESTSAGASTVSGSTGSTVSDSSTISDSTGSTNAPGSTESTVTGSSVSTVSGSTGSTGPST 745

Qy    847 LSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESS---IVTSTVPTASTMSD 903
Db    746 MSASTGSTNTPGSTESTITDGSTVSGSTGSTGSTNNPGSTDSSTTGISTVSGSSLSTISG 805

Qy    904 SLSSTDGISATSSDNVSKSGVSV--TTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE- 960
Db    806 STGST--VSGSSDMTVSTGSTSSPGSTESTVSGASTMSPSTGSSVETSTSGSSVSTVSQS 863

Qy    961 -----------SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLT 1009
Db    864 TSSSTTGQSTVSESSVSTVSSESTISQSTGSTTTGESTVFGSTGSTATGSSTMSASTGST 923

Qy   1010 SEPTNTREQPTTLSTTSNSITEDITTSQPTGD--NGDNTSSTNPVPTVATSTLASASEED 1067
Db    924 DTPGSTES-----TITGSTVTGESTVSGSTGSTITEGSTISESTMTTVGVSTGSTITGES 978

Qy   1068 NKSGSHESASTSLKPSMGEN--SGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTR 1125
Db    979 TVSGSTRSTVT------GESTVSGSTESTVSGSTESTPTVPSTVSGSTGSTVTGEST--- 1029

Qy   1126 EQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVT 1185
Db   1030 ---VSGSTASTSSGSTGSSTEAGSTVSGSSAST-----VTSSTGSSTSGES------TVS 1075

Qy   1186 GGAVASASNDQSHSTSVTNSNSIVSNTPQT-TLSQQVTSSSPST---NTFIASTYDGSGS 1241
Db   1076 GSTVSTVSG--STGSTITGESTVSGSTESTVTAESTVSGSSVSTVSGNT--GSTITGE-S 1130

Qy   1242 IIQHSTWLYGLITLL 1256
Db   1131 TVSGSTGSTGESTIL 1145
```

RESULT 8
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
Submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

```
  Query Match             10.0%;  Score 646.5;  DB 2;  Length 2232;
  Best Local Similarity   23.5%;  Pred. No. 2.3e-20;
  Matches  361;  Conservative  227;  Mismatches  543;  Indels  405;  Gaps   62;

Qy      12 LSIASAKTITGVFDS--FNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTL 69
           || :: :: :| | |   :|  :   : : |     | | : : | :     :  || : |:
Db     274 LSSSTIQSSSGTFPSSVASSPSTVGSTSGAASSSSYATVSTIAGSTGSTITPVPGSSSTI 333

Qy      70 NMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTF-----STLTCTVNDALKSSIKA 124
                 :|  |   :: |    |    ||   :|      |: |   :  : ||
Db     334 G--------SSTPSASSSSSG----TMSTISGSTGSTVTVVPGSSSTFASSTPIASSSSP 381

Qy     125 FGTVTLPIAFNVGGT-GSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 183
             |||  :|     | ||||    |    : : |:: | |    :: |    : |  | :
Db     382 GSTVT--VAPGSSSTYGSSTPSASS----SSSGTMSTNSGSTGSTVTV----APVSSSTF 431

Qy     184 LYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNY 243
             :: :  | :  :|: |     : | | |   |||:   |   |      : |
Db     432 GSSTPIASSSSSGSTVTVVSGSSSTYGSSTPSASSSSAGTASTISG-----STGSTATIV 486

Qy     244 PVSSESF-SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR 302
           | || |  | |:: : :                |   : :|    :  |:   :  : | |
Db     487 PGSSSSVGSSTQSASPSS-------------PGTMSTVSGPTGSTVTVVPGSSTSPAPSS 533

Qy     303 LQSKPFTLRWTGYKNSDAGSNGIVI-----------VATTRTVTDSTTAV----TTLPFN 347
             :   :   ||   : :||: |::                |:::   |:||     :|:| :
Db     534 SPNPSSSPASTGSTITISGSSSIIVSTVSGSTVSGSTGTSQSTLASSTATPGSSSTVPSS 593

Qy     348 PSVDKTKTIEILQPIPTTTITT----------------------SYVGVTTSYLTKTAP 384
           |   :       | | |  ||                        |   :|    | ::|
Db     594 SSPQPSSQ----SPAPNTGSTTPSQTSSQSPSPSMNPSSSTPTGSSQSTITPEGSTASSP 649

Qy     385 IGETATVIVDVPYHTTTTVTSEWT----GTITTTTTRTNP--------TDSIDTV----- 427
           | | :     :   | |||: |     :: | :| ::|        |  : |:
Db     650 TGSTGST-----FSVATEVTSQSTVPSGSSLGTQSTNSSPSPSSLSPSTSGMSTLTSEPS 704

Qy     428 -----------VVQVPLPNPTVSTTEYWSQ-SFATT------TTVTAP------------ 457
                     :  | |||: ||:   |  | |||      ||:|:|
Db     705 PSSTQSSGAQSTLTTPSPNPSQSTSSLESSTSGATTSSGSAGTTMTSPSQSSSVGSSQGS 764

Qy     458 --------------------PG-----------GTDTVIIREPPNHTVTTTEYWSQSFAT 486
                              ||           |   :    |  |||     | | ::
Db     765 TSPAASTTSGEMTSQGSTQTPGSSVSTSAAILTSTQQSVSTNSPGSTVTRPSTVSGSTSS 824

Qy     487 TTTVTAPPGGTD------SVIIREP-------PNPT-------VTTTEYWSQSFATTTTV 526
           :|||   | |:       ||     |       |||:       :| : | ||| :   :
       825 GSTVTV--GSTEASTSGSSVASSSPAPSTSQNPNPSTSSGSSMITQSPYPSQSTSPVESS 882

       527 TAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTE 586
           | |  |:         |  |:|:|   | | : :||: :  | |          |  : |:
Db     883 TTPSPGS--------PGTTLTST---SPSPSQSTTIGSTQGSTS-------PGISTTSEE 924

Qy     587 YWSQSY-----ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGT 641
             ||       :| :||| |   :|:          || :||  |    :|:  | | : :
Db     925 MTSQGSTQTPGSTGSTVTQPSTVSDST--SSGSTVTVGSTEGSSSPIPSTSQNTNPSTSS 982

Qy     642 DTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREP-PNHTVTTTEYWSQSY 700
           : :  : |  : :|:   | :   |:: :  || | | :    | |: |: :    ||
Db     983 GSSMSTQTPQSSQSTSPVESSTSGATSS-SGSPGTTLTSISPSPSPSSTIGS----SQGS 1037

Qy     701 ATTTTVTAPPGETDTVLIREPPNHTVTT-----TEYWSQSYATTTTV-----TAPPGGTD 750
           :    |   | |:|       |  | :|     |   | |  :| |:     ::    ||:
Db    1038 TSPVVSTISQGSTET------PGSTGSTVTKPSTVSGSASSGSTATMGSTEASSTSGGSS 1091

Qy     751 TVIIREPPNPTVTTTEYWS-------QSFATTTTVTAPPGGTDTVIIYESMSSSKISTSS 803
           |      |||: :|:   |         |  | |:::  |  : |:   :  :|  :||:|
Db    1092 T-----SPNPSQSTSPSTSGATSSPGSSGTTLTSISPSPSQSSTIGSSQGSTSPVVSTTS 1146

Qy     804 NDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELI 863
           |:||   |   |    || :  ||  | | :|   |:| |    ||      | :|:  |
Db    1147 GDMTS-QGSTQIPGSTGSTVTQPST-GSGSTSTSGEITSQG----STQTPRSSLSTSPAI 1200

Qy     864 CSDGKECSRLSSSSGIVTNPDSNESSIVT-STVPTAST---------MSDSLSSTDGISA 913
           : ::   :|   || | :   |  : ||| | ||          : ||||:   : :
Db    1201 STSTQQSVSTNSPGSTVTQPSTVRGSTSSGSTVTTGSTEGSSTSGSSSATSLSSSSPVPS 1260

Qy     914 TS-SDNVSKSGVSVTTE---------TSVTTIQ-------TTPNPLSSSVTSLTQLSSIP 956
           || | | | || |  |           | || :         ||: : |:|   || |::
Db    1261 TSQSPNPSTSGSSTPTPNPSQSTSPVVSTTTGEMTSHGSTQTPSTIGSTV---TQPSTVS 1317

Qy     957 SVSESESKVTFTSNGDNQSGTH----------------------DSQSTSTEIEIVTTS 993
             : | | ||  |:  : ||:                        | | ||  :: : |
Db    1318 GSNSSGSTVTIGSSEASTSGSSFKTSPSSISPVPTSSPIPSTTFASSTSGSTISDVSSVS 1377

Qy     994 STKVLP----------------------------PVVSSNTDLTSEPTNTREQPTTLST 1024
           :| : |                            || |  : , : || : |  | ||:
Db    1378 TTSLAPLSSSLPSTVPSSTQSFSSTSEGSSKASSSPVPSQTSSTPTNPTGSTESSTLLSS 1437

Qy    1025 TSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSM 1084
           | :  |:  | |:    :|  : |||   | :| |: |:|     ||   |:::| :| |
Db    1438 TISGSTQHTTMSK--ASSGSTSPSTNS-QTGSTVTMGSSS----TSGVSTSSASSTQPQM 1490

Qy    1085 GENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEP----TDTREQPTTLSTTSKTNSE 1140
             : |  :  : : ::| ||  : :   |:||  :|      :   |  || | :|:| |
Db    1491 STSQGSSAGSTVASSTASPAASSTAPSSTGTMSSTSSGTVGSTISESSTTASASSQTGST 1550

Qy    1141 L---------VATTQATN-------ENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSV 1184
           :          |:|: |::         |  : ||   :| | |  ::::  |   | ||
Db    1551 VTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGST--VASSTAGLVSTSTVPSSTGTMGST 1608

Qy    1185 TGGAV----------ASASNDQ-------SHSTSVTNSNSIVSNTPQTTLSQ-------- 1219
           : | |           ||||:         | |||  :::|  |  || : ||
Db    1609 SSGTVGSTISESSTTASASSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGSTV 1668

Qy    1220 -------QVTSSSPSTNTFIASTYDGS-GSIIQHST 1247
                  ||: ||:   : ||  |: || |  |:
Db    1669 ASSTTGLVSTSTVPSSTGTMGSTSSGTVGSTISESS 1704
```

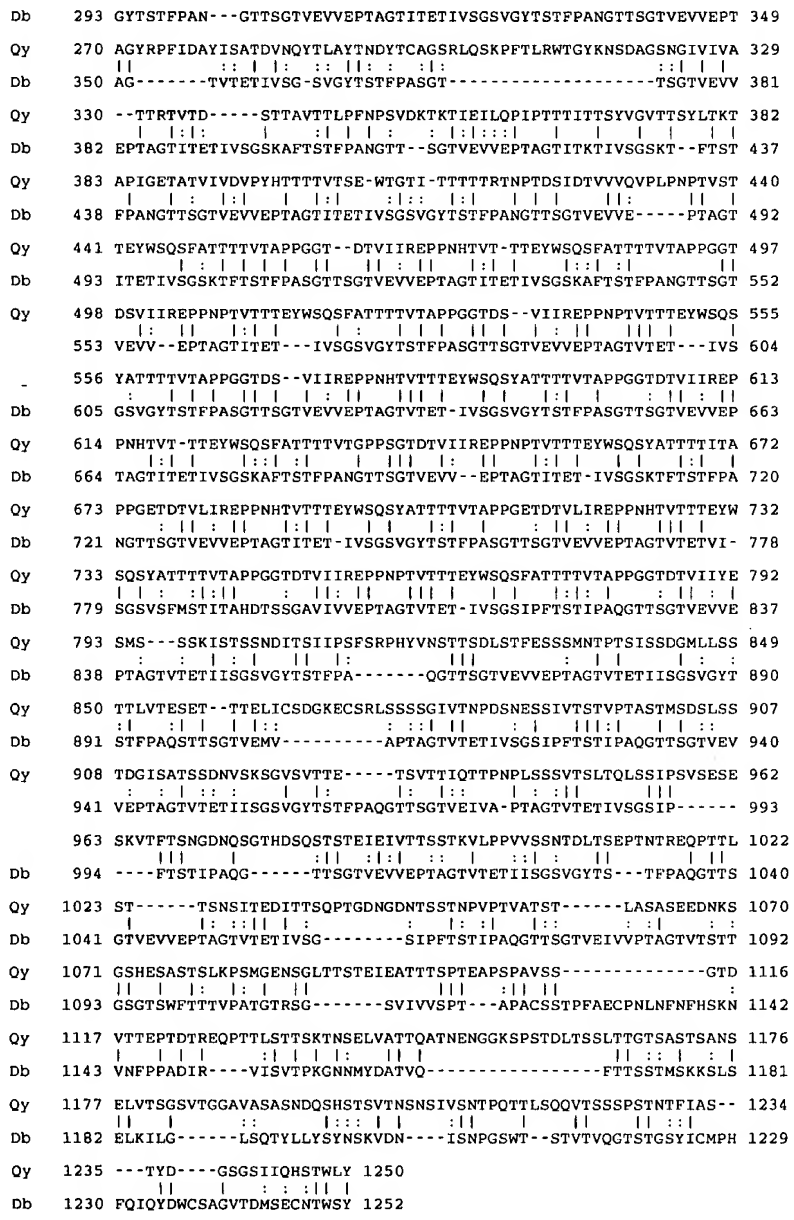
RESULT 9
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oma, A.; Mizutani-Ui,,Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447

```
  Query Match              9.8%;  Score 637.5;  DB 2;  Length 2271;
  Best Local Similarity   22.3%;  Pred. No. 5.9e-20;
  Matches  330;  Conservative  280;  Mismatches  521;  Indels  347;  Gaps   52;

Qy      17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76
           | |:    ||  |  :|   |   |  |:     :|::      :||
Db     297 AITLGTRIDSNKSFHFSGKVNLGNKYEGHGNGGDGIGFAF-----SPG------------ 339

Qy      77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIK--------AFGTV 128
             :| ::  | |:    :  |  | :  |:'  |   | | |::        |||
Db     340 -VLGETGLNGAAVGIGGLSNAF--GFKLDTYHN-TSKPNSAAKANADPSNVAGGGAFGAF 395

Qy     129 TLPIAFNVGGTGSSTDLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186
```

```
Db    396 VTTDSYGVATTYTSSSTADNAAKLNVQPTNN-TFQD------FDINYNGDT--------- 439

Qy    187 SRVMPSLNKVTTLFVAPQ----------CENGYTSGTMGFSSSNGDVAIDCSNIHIGITK 236
Db    440 --------KVMTVKYAGQTWTRNISDWIAKSGTTNFSLSMTASTGG-ATNLQQVQFG--- 487

Qy    237 GLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG--------YRPFIDAYISATDVNQY 288
Db    488 ------------TFEYTESAVT---QVRYVDVTTGKDIIPPKTYSGNVDQVVTIDNQQSA 532

Qy    289 TLAYTNDYTCAGSRLQS---------------KPFTLRWT---------GYKNSDAGS-- 322
Db    533 LTAKGYNYTSVDSSYASTYNDTNKTVKMTNAGQSVTYYFTDVKAPTVTVGNQTIEVGKTM 592

Qy    323 NGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPI------------------PT 364
Db    593 NPVVLTTTDNGTGTVTNTVTGLPSGLSYDSATNSIIGTPTKIGQSTVTVVSTDQANNKST 652

      365 TTITTSYVGVTTSYLTKTAPIGETATVI------------------------------- 392
Db    653 TTFTINVVDTTAPTVT---PIGDQSSEVYSPISPIKIATQDNSGNAVTNTVTGLPSGLTF 709

Qy    393 -----------VDVPYHTTTTVTSEWTGTITTTT-----TRTNPTDSIDTVVVQVPLPNP 436
Db    710 DSTNNTISGTPTNIGTSTISIVSTDASGNKTTTTFKYEVTRNSMSDSVST--SGSTQQSQ 767

Qy    437 TVSTTEYWSQSFATTT------TVTAPPGGTDTVIIRE--PPNHTVTTTEYWSQSFATTT 488
Db    768 SVSTSKADSQSASTSTSGSIVVSTSASTSKSTSVSLSDSVSASKSLSTSESNSVSSSTST 827

Qy    489 ------TVTAPPGG--TDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db    828 SLVNSQSVSSSMSGSVSKSTSLSDSISNSNSTEKSESLSTSTSDSLRTSTSLSDSL---- 883

Qy    541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTV-- 598
Db    884 ----SMSTSGSLSKSQSLSTSISG--SSSTSASLSDSTSNAISTSTSLSESASTSDSISI 937

Qy    599 --------TAPPGGTD----TVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVII 646
Db    938 SNSIANSQSASTSKSDSQSTSISLSTSDSKSMSTSESLSDSTSTSGSVSGSLSIAASQSV 997

Qy    647 REPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT--- 703
Db    998 STSTSDSMSTSEIVSDSISTSGSLSASDSKSMSVSSSMSTSQSGSTSESLSDSQSTSDSD 1057

      704 ----TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVII----- 754
Db   1058 SKSLSLSTSQSGSTST-------STSTSASVRTSESQSTSGSMSA--SQSDSMSISTSFS 1108

Qy    755 ---REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIP 811
Db   1109 DSTSDSKSASTASSESISQSASTSTS-----GSVSTSTSLSTSNSERTSTSVSDSTSLST 1163

Qy    812 SFSRPHYVNSTTSDLSTFESSSMNTPTSIS-------SDGMLLSSTTLVTE------SET 858
Db   1164 SESDSISESTSTSD-SISEAISASESTSISLSESNSTSDSESQSASAFLSESLSESTSES 1222

Qy    859 TTELICSDGKECSRLS-------SSSGIVTNPDSNESSIVTSTVPTAST----------- 900
Db   1223 TSESVSSSTSESTSLSDSTSESGSTSTSLSNSTSGSASISTSTSISESTSTFKSESVSTS 1282

Qy    901 ----MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIP 956
Db   1283 LSMSTSTSLSNSTSLSTSLSDSTSDS-KSDSLSTSMSTSDSISTSKSDSISTSTSLSGST 1341

Qy    957 SVSESESKVTFTSNGDNQS-------GTHDSQSTSTEIEIVTTSSTKV------------ 997
Db   1342 SESESDSTSSSESKSDSTSMSISMSQSTSGSTSTSTSTSLSDSTSTSLSLSASMNQSGVD 1401

Qy    998 --LPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTV 1055

Db   1402 SNSASQSASNSTSTSTSESDSQSTSTYTSQSTSQSESTSTSTSLSDSTSISKSTSQSGST 1461

Qy   1056 ATSTLASASEEDNKSGS---------HESASTSLKPSMG-ENSG-LTTSTEI-EATTTSP 1103
Db   1462 STSASLSGSESESDSQSISTSASESTSESASTSLSDSTSTSNSGSASTSTSLSNSASASE 1521

Qy   1104 TEAPSPAVSSGTDVTTEPTDTREQPT------TLSTTSKTNSELVA---TTQATNENGGK 1154
Db   1522 SDSSSTSLSDSTSASMQSSESDSQSTSASLSDSLSTSTSNRMSTIASLSTSVSTSESGST 1581

Qy   1155 SPSTDLTSSLTT------GTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSI 1208
Db   1582 SESTSESDSTSTSLSDSQSTSRSTSASGSASTSTSTSDSRSTSASTSTSMRTSTSDSQSM 1641

Qy   1209 -VSNTPQTTLSQQVT---SSSPSTNTFIASTYDGSGSI 1242
Db   1642 SLSTSTSTSMSDSTSLSDSVSDSTSDSTSASTSGSMSV 1679


RESULT 10
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000.#sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
A;Reference number: Z22899; MUID:97166151
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B


  Query Match             9.7%;  Score 632.5;  DB 2;  Length 3570;
  Best Local Similarity   24.9%;  Pred. No. 1.7e-19;
  Matches  340;  Conservative  155;  Mismatches  509;  Indels  363;  Gaps   61;

Qy      60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQF-YSGEEFTTFSTLTCTVNDAL 118
Db    2329 SSTPGTTWILT-----KLTTTATTTESTGSTATPSSTQGPPAGTPHVSTTATTPTVTSSK 2383

Qy     119 KSSIKAFGTVT-LPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKST 177
Db    2384 ATPFSSPGTATALPALRSTATTPTATSFTAIPSSSLGTTWTRLSQTTTPMA-----TMST 2438

Qy     178 VDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKG 237
Db    2439 ATPSS-------TPETVHTSTVLTTTATTTGATGSVATPSSTPG--TAHTTKVPTTTTTG 2489

Qy     238 LNDWNYPVSSES--------FSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYT 289
Db    2490 FT--VTPSSSPGTARTPPVWISTTTTPTTSGSTVTPSSIP-----------GTTHTPTV 2535

Qy     290 LAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPS 349
Db    2536 LTTTTQPVATGSMATPSSST-------QTSGTPPSLITTATTITATGSTT-------NPS 2581

Qy     350 VDKTKTIEILQPIP---TTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTT----- 401
Db    2582 STPGTT-----PIPPELTTTATTP--AATSSTVTPSSALGTTHT--PPVPNTTATTHGRS 2632

Qy     402 -------TVTSEWT----GTITT-------TTTRTNPTDSIDTVVVQVPL---PNPTVST 440
Db    2633 LSPSSPHTVRTAWTSATSGTLGTTHITEPSTGTSHTPAATTGTTTTSTPALSSPHPSSRT 2692

Qy     441 TEYWSQSFATT---------TTVTAPPG--GTDTVIIREP---PNHTVTTTE-----YWS 481
```

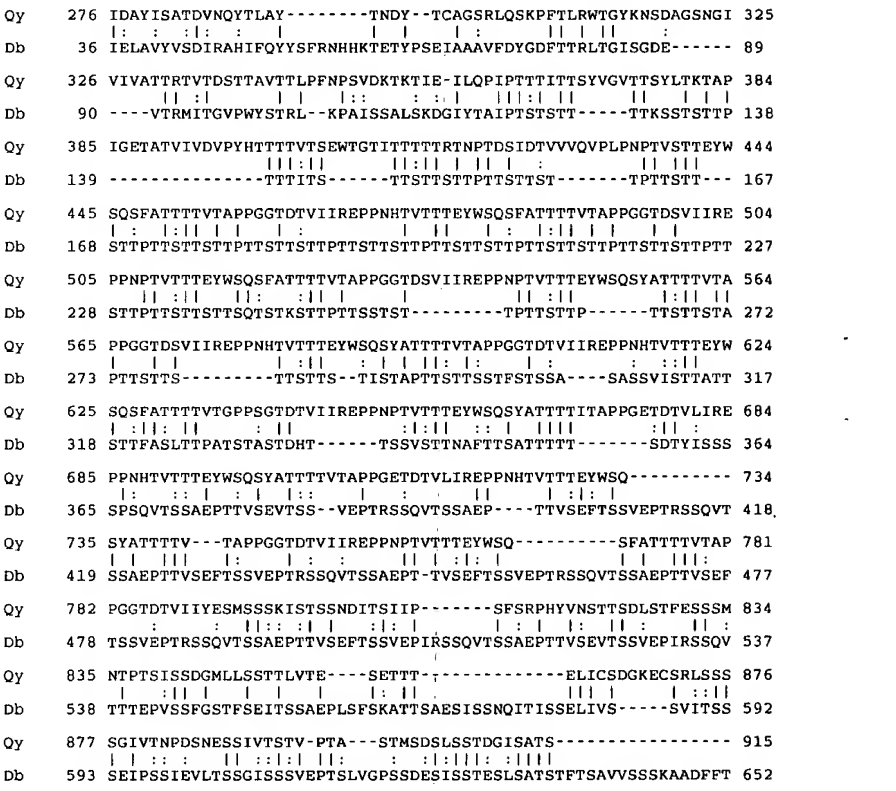
```
Db    2693 TESPPSPGTTTPGHTTATSRTTATATPSKTRTSTLLPSQPTSAPITTVVTTGCEPQCAWS 2752

Qy     482 Q--------------SFATTTTV------------------TAPPG------------GT 497
           :                | | : :                  || ||
Db    2753 EWLDYSYPMPGPSGGDFDTYSNIRAAGGAVCEQPLGLECRATAQPGVPLGELGQVVECSL 2812

Qy     498 DSVII---RE-----------------------PPNPTVTTTEYWSQS---------FAT 522
           |  ::   ||                       |  |  ::|   | :           |
Db    2813 DFGLVCRNREQVGKFKMCFNYEIRVFCCNYGHCPSTPATSSTAMPSSTPGTTWILTELTT 2872

Qy     523 TTTVTAPPGGT------------DSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTD 570
           | | ||  | |               |:      || |:::  | |   | | | |
Db    2873 TATTTASTGSTATPSSTPGTAPPPKVLTSPATTPTATSSKATSSSSPRTAT-TLP----- 2926

Qy     571 SVIIREPPNHTVTT-TEYWSQSYATTTTV----TAPPGGTDTVIIREPPNHTVTTTEYWS 625
           |:       | |: |   | :  || |:    | |      |:     |  | |:|   :
Db    2927 -VLTSTATKSTATSVTPIPSSTLGTTGTLPEQTTTPVATMSTIHPSSTPETTHTSTVLTT 2985

Qy     626 QSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREP 685
           ::    |: |  || |        |  |   ||       |  | ||  | | |
Db    2986 KATTRATSSTSTPSST--------PGTTWILTE-----LTTAATTTAGTGPTAT------ 3026

Qy     686 PNHTVTTTEYWSQSYATTT-TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
           |: |  ||  |  :  ||| | ||  | | |  :   |  |   ||      :|| ||||
Db    3027 PSSTPGTT--WILTELTTTATTTASTGSTAT--LSSTPGTTWILTE-----PSTTATVTA 3077

Qy     745 PPGGTDTVIIREPP--NPTVTTTEYWSQSFATTTTVTA----PPGGTDTVIIYESMSSSK 798
           ||| | |    :     | |:||       ||| |||:     |      |      :: |:
Db    3078 PPGSTATASSTQATAGTPHVSTT-------ATTPTVTSSKATPSSSPGTATALPALRSTA 3130

Qy     799 ISTSSNDITSIIPSFSRPHYVNSTTSDLS--TFESSSMNTPTSISSDGMLLSSTTLVTES 856
           : ::   |: ||| |    : :| : ||  |  :::|:| | |:   : :|| | | :
Db    3131 TTPTATSFTA-IPSSS----LGTTWTRLSQTTTPTATMSTATPSSTPETVHTSTVLTTTA 3185

Qy     857 ETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSS 916
           ||                 ::| |  | |   :  |: ||| '       ::| | :|| |
Db    3186 TTT---------------GATGSVATPSSTPGTAHTTKVPT--------TTTGFTATPS 3222

Qy     917 DNVSKSGVSVT----TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD 972
             | | ::|      |: |   |||    |:||     ||||
Db    3223 ---SSPGTALTPPVWISTTTTPTTTTPTTSGSTVTP----SSIPGT-------------- 3261

Qy     973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITED 1032
             || ::      ::||::|    | : :   |  | |   | :|:||: :||
Db    3262 ----THTAR-------VLTTTTT----TVATGSMATPSSSTQTSGTPPSLTTTATTITAT 3306

Qy    1033 ITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTT 1092
           :|: |:     :|  | |:| | ||   : :   :|      |::|  |       : |
      3307 GSTTNPS-----STPGTTPIPPVLTSMATTPAATSSK------ATSSSSPRTATTLPVLT 3355

      1093 STEIEATTTSPTEAPSPAVSSGTDV---TTEPTDTREQPTTLSTTSKTNSELVATTQATN 1149
           ||  ::| || |  ||  : :   |   || |  |      | ||   |::  | || ||
Db    3356 STATKSTATSFTPIPSSTLWTTWTVPAQTTTPMSTMSTIHTSSTPETTHTSTVLTTTATM 3415

Qy    1150 ENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTS-------- 1201
             :: ::  | | | ||   |   :|| |: : |    ::|:   :  |:
Db    3416 T---RATNSTATPSSTLGT---TRILTELTTTATTTAATGSTATLSSTPGTTWILTEPST 3469

Qy    1202 ------VTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSI 1242
                 | | :  |:|  |  : :| ::  :  |  |||   | ::
Db    3470 IATVMVPTGSTATTSSTLGTAHTPKVVTAMATMPTATASTVPSSSTV 3516


RESULT 11
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A;Reference number: A45106; MUID:93016075
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A;Reference number: A43932; MUID:91358717
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; v
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>


  Query Match             9.2%;  Score 599;  DB 2;  Length 3020;
  Best Local Similarity   25.8%;  Pred. No. 3.9e-18;
```

```
  Matches  285;  Conservative  120;  Mismatches  438;  Indels  262;  Gaps   44;

Qy     249 SFSYTKTCTSNGIQIKYQNV------PAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR 302
           :| | : |  ||    |: |:      |:    |    :  |  : :|    |    : :
Db    1236 AFCYWEICGPNGTVEKHFNICSITTRPSTLTTFTTITLPTTPTS-FTTTTTTTPTSSTV 1294

Qy     303 LQSKP-FTLRWTGYKNSDAGSNGI----------VIVA----TTRTVTDSTTAVT----- 342
           | : |     |: : | |  |:|           |  |       |:| |   ::
Db    1295 LSTTPKLCCLWSDWINEDHPSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQK 1354

Qy     343 ------------------TLPFNPSVDKTKTIE--------ILQPIPTTTITTSYVGVTT 376
                             ||    |    :         |  | | ||  :     ||
Db    1355 VQCDVSVGFICKNEDQFGNGPFGLCYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTT 1414

Qy     377 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTT---RTNPTDSIDTVVVQVPL 433
           :    | |   | |         ||   |:  :  ||||||    | |:  | |     |
Db    1415 TLPPTTTPSPPTTTT-------TTPPPTTTPSPPITTTTTPLPTTTPSPPISTTTTPPPT 1467

Qy     434 PNPTVSTT----EYWSQSFATTTTVTAPPGGTDTVIIR---EPPNHTVTTTEYWSQSFAT 486
             |:  ||       : |  |||| | ||  | :  :     ||  | |     : |  |
Db    1468 TTPSPPTTTPSPPTTTPSPPTTTTTTPPPTTTPSPPMTTPITPPASTTTLPPTTTPSPPT 1527

Qy     487 TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTV 546
           ||| | ||  | |      ||  |  |    : :   ||| : ||  | :       |:|
Db    1528 TTTTTPPPTTTPS-----PPTTTPITPPTSTTTLPPTTTPSPPPTTTTPPPTTTPSPPT 1582

Qy     547 TTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD 606
           |||      |  | || | ||  | |      ||  | ||     :      |||    ||  |
Db    1583 TTT----PSPPTITTTTPPPTTTPS-----PPTTTTTTPPPTTTPSPPTTTPITPPTSTT 1633

Qy     607 TVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYAT 666
           |:    ||    |||      |   ||| | ||: |         |:|  |||     |
Db    1634 TL----PP----TTT----PSPPPTTTTPPPTTT--------PSPPTTTT----PSPPI 1669

Qy     667 TTTITAPPGETDTVLIR---EPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPN 723
           ||| | ||  | :  |     ||  |:||     :   :  || | |   |        ||
Db    1670 TTTTTPPPTTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTP----SPPP 1725

Qy     724 HTVTT-TEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAP- 781
            |:|| :   : |  |||  | ||  | : :   |  |::|    :   |  | ||    |
Db    1726 TTMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPL 1785

Qy     782 ----------------PGGTDTVII------------------YESMSSSKI-STSSNDI 806
                           ||| || :|                  |  :   ::  |   |:
Db    1786 CNWTGWLDSGKPNFHKPGG-DTELIGDVCGPGWAANISCRATMYPDVPIGQLGQTVVCDV 1844

Qy     807 T-------------SIIP-SFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTL 852
           :              :|| :|   : :|    :  | : ::| | |: :     :::||
Db    1845 SVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVT-QPTTMTTTTTENPTPTPITTTTT 1903

Qy     853 VTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGIS 912
           || : | |     :| :              |     |  |:  || :       :|  |:
Db    1904 VTPTPTPTSTQSPNGLQA------------PTPTPISTTTTVTPTPTPTGTQTPTTTPIT 1951

Qy     913 ATSSDNVSKSGVSVTTETSV-----TTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTF 967
             |::   : :      | |:|     ||:  || | |:  |::| :::  :|: : :
Db    1952 TTTTVTPTPTPTGTQTPTTVLITTTTTMTPTPTPTSTKSTTVTPITTTTTVTATPT---- 2007

Qy     968 TSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSN 1027
                 :||     | | | ||::   : |  :  |  ||:|    |    ||||
Db    2008 ------PTGT----QTPTMIPISTTTT---VTPTPTPTTGSTGPPTHTSTAPIAELTTSN 2054

Qy    1028 SITEDIT------TSQPTGDNG----------DNTSSTNP-VPTVATSTLASASEEDNKS 1070
             |  |       || |  ::           : ||:  |  ||  |:|            |
Db    2055 PPPESSTPQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPPSTPTAPTTT----------S 2104

Qy    1071 GSHESA---STSLKP--------SMGENSGLTTSTEIEATTTS-----PTEAPSPAVSSG 1114
           | |  :   ||:  |         : | :|  | || :: ||||      ||   :|::
Db    2105 GGHTLSPPPSTTTSPPGTPTRGTTTGSSSAPTPST-VQTTTTSAWTPTPTPLSTPSIIRT 2163

Qy    1115 TDVTTEPT---------DTREQP-------TTLSTTSKTNSELVATTQATNENGGKSPST 1158
           | :   |:          ||  |       |   |    |  |  | :  | :   :||
Db    2164 TGLRPYPSSVLICCVLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNWSCPSTPSP 2223

Qy    1159 DLTSSLTTGTSASTSANSELVTSGS 1183
             | | :| | :  |:      | |:
Db    2224 TPTPSKSTPTPSKPSSTPSKPTPGT 2248


RESULT  12
T39174
hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (S
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50375; T39172; T39173; T39174; T39366
R;Connor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21832
A;Accession: T50375
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1283 <CON>
A;Cross-references: EMBL:AL021815; PIDN:CAB61533.1; GSPDB:GN00067; SPDB:SPBC8E4.70c
A;Experimental source: strain 972h-; cosmid c8E4
A;Accession: T39172
A;Molecule type: DNA
A;Residues: 785-1283 <CO2>
A;Cross-references: EMBL:AL021815; PIDN:CAA17000.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
A;Accession: T39173
A;Molecule type: DNA
A;Residues: 'ME',179,'PLV',183,'W','KL',556-761,'HRGSS' <CO3>
A;Cross-references: EMBL:AL021815; PIDN:CAA17001.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
A;Accession: T39174
A;Molecule type: DNA
A;Residues: 1-555,'S' <CO4>
A;Cross-references: EMBL:AL021815; PIDN:CAA17002.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21848
A;Accession: T39366
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-896,'G',898-904,'I',906-1283 <WOO>
A;Cross-references: EMBL:AL035675; PIDN:CAB38695.1; GSPDB:GN00067; SPDB:SPBC1289.15
A;Experimental source: strain 972h-; cosmid c1289
C;Genetics:
A;Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07c
A;Map position: 2
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida


  Query Match             8.9%;  Score 578;  DB 2;  Length 1283;
  Best Local Similarity   25.6%;  Pred. No. 1.1e-17;
  Matches  344;  Conservative  170;  Mismatches  571;  Indels  258;  Gaps   58;

Qy      13 SIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMP 72
           | ::|   | :    ||    | :         |  :      |    :| | :| |  :
Db      63 SFSNASVYTTLITITNSQCTGNYSTLTIS--SQPLIHTNTSISKPSQTATPQNTNTTQVS 120

Qy      73 CVFKYTTSQTS------VDLTADGVK---YATCQFYSGEEFTTFSTLTCTV--NDALKSS 121
             ||:  |      :  | :|       |  : :      :|| | |   |  :  |
Db     121 LTNGTTTNSYSNTNSLPITDTINGTTELIIPTTSYNNQSHTLIYSTYTSTYLPNSTIDLS 180

Qy     122 IKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKS----- 176
           |    |::    ::  | :|  |  :  ||||  | |    |   :      |
Db     181 ILPHSTISTLSTVSINDTSAS--LSKTTSPTAGTITETIVSG--SVGYTSTFPASGTTSG 236

Qy     177 ---TVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMG------------FSSSN 220
               |:|:|      ::      |: | | |    || ||||:              | |
Db     237 TVEVVEPTAGTITETIVSGSVGYTSTFPA----NGTTSGTVEVVEPTAGTITETIVSGSV 292

Qy     221 GDVAIDCSNIHIGITKGLNDWNYP--------VSSESFSYTKTCTSNGI---QIKYQNVP 269
           |  :   :|   | | |  :   |         : | |  || |  :||     ::
```



RESULT 13
S57180
probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J2223; serine/threonine-rich protein YJR151c
C;Species: Saccharomyces cerevisiae
C;Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 05-Nov-1999
C;Accession: S57180
R;Scarcez, T.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57169
A;Accession: S57180
A;Molecule type: DNA
A;Residues: 1-1161 <SCA>
A;Cross-references: EMBL:Z49651; NID:g1015902; PID:g1015903; GSPDB:GN00010; MIPS:YJR1
C;Genetics:
A;Gene: MIPS:YJR151c
A;Map position: 10R
C;Keywords: transmembrane protein

Query Match 8.9%; Score 577; DB 2; Length 1161;
Best Local Similarity 26.3%; Pred. No. 1e-17;
Matches 293; Conservative 131; Mismatches 427; Indels 264; Gaps 46;



```
Qy     916 -SDNVSKSGVSVTTETSVTTIQTTPN-PLSSSVTSLTQ---------LSSIPSVSESES 963
Db     653 RSTVSAKSDVSGNSSTQSTTFFATPSTPLAVSSTVVTSSTDSVSPNIPFSEISSSPESST 712

Qy     964 KVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNT-----REQ 1018
Db     713 AITSTSTSFIAERTSSLYLSSSNMSSFTLSTFTVSQSIVSS---FSMEPTSSVASFASSS 769

Qy    1019 PTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTV-----ATSTLASASEEDNKSGSH 1073
Db     770 PLLVSSRSN--CSDARSS--------NTISSGLFSTIENVRNATSTFTNLSTD------- 812

Qy    1074 ESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTT--EPTDTREQPTTL 1131
Db     813 EIVITSCKSSCTNEDSVLTKTQVSTVETTITSC------SGGICTTLMSPVTTINAKANT 866

Qy    1132 STTSKTNSELVATTQATNENGGKS----PSTDLTSSLTTGTSASTSANSELVTS------ 1181
Db     867 LTTTETST--VETTITTCPGGVCSTLTVPVTTITSEATTATISCEDNEEDITSTETELL 924

Qy    1182 ------GSVTGGAVASASN-----DQSHSTSVTNSNSIVSNTPQT-------TLSQQVTS 1223
Db     925 TLETTITSCSGGICTTLMSPVTTINAKANTLTTTETSTVETTITTCSGGVCSTLTVPVTT 984

Qy    1224 -SSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLS 1257
Db     985 ITSEATTATISCEDNEEDVASTKTELLTMETTIT 1019
```

RESULT 14
T32271
hypothetical protein ZC178.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32271
R;Murray, J.; Wohldmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid ZC178.
A;Reference number: Z21143
A;Accession: T32271
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1459 <MUR>
A;Cross-references: EMBL:AF024496; PIDN:AAB70340.1; GSPDB:GN00023; CESP:ZC178.2
A;Experimental source: strain Bristol N2; clone ZC178
C;Genetics:
A;Gene: CESP:ZC178.2
A;Map position: 5
A;Introns: 1099/2

```
  Query Match             8.5%;  Score 553.5;  DB 2;  Length 1459;
  Best Local Similarity   24.5%;  Pred. No. 1.4e-16;
  Matches  333;  Conservative  172;  Mismatches  593;  Indels  261;  Gaps   55;

Qy      57 DGTSANPGDTFT----LNMPCVFKYTTSQTSVDLTADGVKYAT----------------- 95
Db     158 DSTTAGPSSTATNSASSETPC-----NSETQ---TSDGTSTMTVPNDSTTAGPSSTATNS 209

Qy      96 ------CQFYSGEEFTTFSTLTCTV-NDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDS 148
Db     210 ASSETPC---NSETQTSDGTSTMTVPNDS---------TTAGPSSTATNSASSETPCNSE 257

Qy     149 KCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENG 208
Db     258 TQTSDGTSTMT----------VPNDSTTAGPS-----STATNSASSETPCNSETQTSDG 301

Qy     209 YTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNG---IQIKY 265
Db     302 TSTMTVPNDSTTAGPSSTATN--------------SASSETPCNSETQTSDGTSTMTVPN 347

Qy     266 QNVPAGYRPFIDAYISA----------TDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTG 314
Db     348 DSTTAG--PSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTATNSASSETP 405

Qy     315 YKNSDAGSNGIVIVATTRTV-TDSTTA-----VTT-----LPFNPSV---DKTKTIEI-- 358
Db     406 CNSETQTSDG----TSTMTVPNDSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVSS 461

Qy     359 --LQPIPTTTITTSYVGVT-------TSYLTKTAPI-------GETATVIVDVPYHT--- 399
Db     462 DSTTPGPSSTATNSASSETPCNSETQTSDGTFTMTVSSDSTTAGPSSTVTNSASSETPCN 521

Qy     400 TTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAP-- 457
Db     522 SETQTSDGTSTMTVSSDSTTPGPS-STATNSASSETPCNSET----QTSDGTSTMTVPND 576

Qy     458 -----PGGTDT--VIIREPPNHTVTTTEYWSQSFATTTTVTAPPGG--TDSVIIREPPNP 508
Db     577 STTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTVTNSASSETPCNS 636

Qy     509 TVTTTEYWSQSFATTTTVTAPPG--GTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPP 566
Db     637 ETQTSDGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVSSDSTTAGP 696

Qy     567 G--GTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDT-VIIREPP--NHTVTTT 621
Db     697 SSTATNSASSETPCNSETQTSDGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETQTSD 756

Qy     622 EYWSQSFATTTTVTGPPSG-TDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTV 680
Db     757 GTFTMTVPNDSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVSNDSTTAGPSSTAT- 815

Qy     681 LIREPPNHTVTTTEYWSQSYATTTTV-----TAPPGETDT--VLIREPPNHTVTTTEYWS 733
Db     816 --NSASSETPCNSETQTSDGTFTMTVPNDSTTAGPSSTATNSASSETPCNSETQTSDGTS 873

Qy     734 QSYATTTTVTAPPGGTDT--VIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGT------ 785
Db     874 TMTVPNDSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTATNSAS 933

Qy     786 -DTVIIYESMSSSKISTSS--NDITSIIPSFSRPHYVNSTT---SDLSTFESSSMNTPTS 839
Db     934 SETPCNSETQTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPN 993

Qy     840 ISSDGMLLSSTTLVTESET--TTELICSDGKEC------SRLSSSSGIVTNPDSNES--S 889
Db     994 DSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTVTNSASSETPCN 1053

Qy     890 IVTSTVPTASTMS-DSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTS 948
Db    1054 SETQTSDGTSTMTVSSDSTTPGPSSTATNSASCETPFVYSHKFCKYDSTTAGPSSTATNS 1113

Qy     949 LTQLSSIPSVSESE-------------------SKVTFTSNGDNQSGTHDSQSTSTEIE 988
Db    1114 AS--SETPCNSETQTSDGTSTMTVPNDSTTAGPSSTVTNSASSETPCNSETQTSDGTFTM 1171

Qy     989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048
Db    1172 TVSSDSTTAGPSSTVTNSASSETPCNSETQ-TSDGTFTMTVSSDSTTAGPSSTVTNSASS 1230

Qy    1049 TNP---------------VPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
Db    1231 ETPCNSETQTSDGTSTMTVPN--DSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVS 1288

Qy    1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNENGG 1153
Db    1289 S--DSTTAGPSSTVTNSASSETPCNSE-TQTSDGTSTMTVSSDSTTAGQSST-ATNSASS 1344

Qy    1154 KSP---------------STDLT-----SSLTTGTSASTSANSELVTSGSVTGGAVASA 1192
Db    1345 ETPCNSETQTSDGTSTMTVSSDSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVSSD 1404

Qy    1193 SNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTF 1231
```

```
Db   1405 STTAGQSSTATNSAS--SETPCNSETQ--TSDKTSTMTF 1439


RESULT  15
S25345
probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YCR1102
C;Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C;Accession: S25345; S19504
R;Wilson, C.; Grisanti, P.; Frontali, L.
Yeast 8, 569-575, 1992
A;Title: The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromos
A;Reference number: S25345; MUID:92397594
A;Accession: S25345
A;Molecule type: DNA
A;Residues: 1-1609 <WIL>
A;Cross-references: GB:X59720; EMBL:S43845; NID:g1907116; PIDN:CAA42254.1; PID:g1907227
R;Frontali, L.; Grisanti, P.
Submitted to the Protein Sequence Database, March 1992
A;Reference number: S19504
A;Accession: S19504
A;Molecule type: DNA
A;Residues: 1-1609 <FRO>
A;Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089w
C;Genetics:
A;Gene: SGD:FIG2
A;Cross-references: SGD:S0000685; MIPS:YCR089w
A;Map position: 3R
C;Keywords: transmembrane protein
F;4-20/Domain: transmembrane #status predicted <TM1>
F;1592-1609/Domain: transmembrane #status predicted <TM2>


  Query Match             8.4%;  Score 547.5;  DB 2;  Length 1609;
  Best Local Similarity   22.5%;  Pred. No. 3e-16;
  Matches  332;  Conservative  232;  Mismatches  503;  Indels  411;  Gaps   66;

Qy      55 SLDGTSANPGDTFT-LNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCT 113
           ||  || ||  ::|  |     | :   :::| |:  :  :|    | | | |  |:| :
Db     165 SLTFTSVNPSQSWTSFNSE---KSSALSSTIDFTSSEISGSTSP-KSLESFDTTGTITSS 220

Qy     114 VNDA-------------------------LKSSIKAF----GTVTLPIAFNVGGTGSST 143
             : :                        | |:|:|      : |:|   :   :   |
Db     221 YSPSPSSKNSNQTSLLSPLEPLSSSSGDLILSSTIQATTNDQTSKTIPTLVDATSSLPPT 280

Qy     144 DLEDSKCFTAGTNTVTFN--------DGDKDISIDVEFEKSTVDPSAYL----YASRVMP 191
               |  |:|:::::: |          |:  |:          :::|||  :     |:|  :
       281 LRSSSMAPTSGSDSISHNFTSPPSKTSGNYDV-----LTSNSIDPSLFTTTSEYSSTQLS 335

       192 SLNKVT--------------------TLFVAPQCENGYT-SGTMGFSSSNGDVAIDCSN 229
           |||: :                      |  : |    | | |  :| |::|     : :
Db     336 SLNRASKSETVNFTASIASTPFGTDSATSLIDPISSVGSTASSFVGISTANFSTQGNSNY 395

Qy     230 IHIGITKG---LNDW---NYPVS--------------------SESFSYTKTCTSNGIQI 263
           :      |     ||   : |:|                    | ::  | | | :|:
Db     396 VPESTASGSSQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVSTATKTVDGVIT 455

Qy     264 KY---------QNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTG 314
           :|        ::   |    | :   |:  :  ::  :|  | | |    : | :
Db     456 EYVTWCPLTQTKSQAIGVSSSISSVPQASSFSGSSILSSNSSTLAAS--NNVPESTASGS 513

Qy     315 YKNSDAGSNGIVIVATT-----RTVTDSTTAVTTLPFNPSVDKTKTI-----EILQPIPT 364
           :  |  |: : :  ||      | |  :   || |   |   |||:     | :   |
Db     514 SQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVST-ATKTVDGVITEYVTWCPL 572

Qy     365 TTITTSYVGVTTSYLTKT--------APIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
           |   :  :|:::| :: |          :| :   : |  :  | |:    |  :| :|:
Db     573 TQTKSQAIGISSSTISATQTSKPSSILTLGISTLQLSDATFKGTETIN---THLMTESTS 629

Qy     417 RTNPT------DSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPP 470
           | ||      ||      :| | : ::|:   :| |  :| |:|
Db     630 ITEPTYFSGTSDSFYLCTSEVNLAS-SLSSYPNFSSSEGSTATIT--------------- 673
```

```
Qy     471 NHTVT---TTEYWSQSFATTTTVTA-----------PPGGTDSVIIREPPNPTVTTTEY 515
           | |||   |::| | | :  |  :                    |::: :    : | :  :|
Db     674 NSTVTFGSTSKYPSTSVSNPTEASQHVSSSVNSLTDFTSNSTETIAVISNIHKTSSNKDY 733

Qy     516 WSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT------TEY--W--SQSYATTTTVTAP 565
             : ||      |  ::::      ||||      ||| |  : | | ||:::
Db     734 ------SLTTTQLKTSGKQTLVL-----STVTTTVNGAATEYTTWCPASSIAYTTSISY- 781

Qy     566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTV---TTTE 622
                         | |||  | |  | | :|:    : |: :    : ||   | :|
Db     782 ------------KTLVLTTEVCSHSECTPTVITSVTATSSTIPLLSTSSSTVLSSTVSE 828

Qy     623 YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI 682
                 |:  |:    | |           | |:|:       : |: ||   |: |
Db     829 GAKNPAASEVTINTQVSATS--------EATSTSTQ------VSATSATATASESSTTSQ 874

Qy     683 REPPNHTVTTTEYWSQSYATTTTVTAPPGETD----TVLIREPPNHTVTTTEYWSQSYAT 738
             : |::|    :|:: || ::  |   |:    ||: |:     : :||  | |
Db     875 VSTASETIST--LGTQNFTTTGSLLFPALSTEMINTTVVSRKT---LIISTEVCSHSKCV 929

Qy     739 TTTVT-------APPGG-------TDTVIIREPPNPTVT-TTEYWSQSFATTTTVTAPPG 783
             | :|        |  |       |: | :    : ||| :||  | |  | | :|:
Db     930 PTVITEVVTSKGTPSNGHSSQTLQTEAVEVTLSSHQTVTMSTEVCSNSICTPTVITSVQ- 988

Qy     784 GTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSD 843
             |   | : |:|  | :|   :|:            :|::|||  |: : | : :
Db     989 MRSTPFPYLTSSTSSSSLASTKKSSL----------EASSEMSTFSVSTQSLPLAFTCS 1037

Qy     844 GMLLSSTTLVTESETT--TELICSDGKECSRLSSSSGIVTN--PDSNES--------SIV 891
               ||| |::   |  |  | |       |||: | ||  | |  |           |:
Db    1038 EK--RSTTSVSQWSNTVLTNTIMS--------SSSNVISTNEKPSSTTSPYNFSSGYSLP 1087

Qy     892 TSTVPTASTMSDSLSSTDGI-----------------------------SATSSDNVS 920
           :|: |:  ::| : :: :||                               |:  | ::|
Db    1088 SSSTPSQYSLSTATTTINGIKTVYTTWCPLAEKSTVAASSQSSRSVDRFVSSSKPSSSLS 1147

Qy     921 KSGVSVTTETSVTTIQ------TTPNPLSSSVT--SLTQLSSIPSVSESESKVTFTSNGD 972
           :: :  |  |: |||        ||  ||:|  |  : || |        | :||  ||
Db    1148 QTSIQYTLSTATTTISGLKTVYTTWCPLTSKSTLGATTQTS-------STAKVRITS-AS 1199

Qy     973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE-PTNTREQPTTLS-----TTS 1026
           : : |  | |||||  |    ||:  |   | | |: | : || :    :||:      :||
Db    1200 SATSTSISLSTSTESE----SSSGYLSKGVCSGTECTQDVPTQSSSPASTLAYSPSVSTS 1255

Qy    1027 NSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGE 1086
           :| :   ||:          ||:   || : :|:  |||   : |      ||||
Db    1256 SSSSFSTTTASTL------TSTHTSVPLLPSSSSISASSPSSTS----LLSTSLPSPAFT 1305

Qy    1087 NSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSEL----- 1141
           :| | |:| : ::|   :  |   :|| : ::  |  :    :  |::| ::| |
Db    1306 SSTLPTATAVSSSTFIASSLP---LSSKSSLSLSPVSSSILMSQFSSSSSSSSSLASLPS 1362

Qy    1142 ------------------VATTQATNENGGKSPST--------DLTSSLT------TGTS 1169
                           :||   |:   :  ||         |:||: |      | |
Db    1363 LSISPTVDTVSVLQPTTSIATLTCTDSQCQQEVSTICNGSNCDDVTSTATTPPSTVTDTM 1422

Qy    1170 ASTSANSELVTSGSVTG-----------GAVASASND--QSHSTSVTNS-----NSIVS 1210
           | :  :  || |  |                ::: | :  |: :||  ||      |:::
Db    1423 TCTGSECQKTTSSSCDGYSCKVSETYKSSATISACSGEGCQASATSELNSQYVTMTSVIT 1482

Qy    1211 NTPQTTLSQQV--TSSSPSTNTFIASTY---DGSGSII 1243
           :  || | :|  | |: |  |    ||   | :| :|
Db    1483 PSAITTTSVEVHSTESTISITTVKPVTYTSSDTNGELI 1520


Search completed: October  3, 2002, 15:33:15
Job time: 252 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:32:34 ; Search time 18.04 Seconds
(without alignments)
2704.359 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6465	99.5	1260	1	ALS1_CANAL	P46590 candida alb
2	4204.5	64.7	1119	1	ALS3_CANAL	O74623 candida alb
3	3194	49.2	1419	1	ALA1_CANAL	O13368 candida alb
4	1992.5	30.7	468	1	ALS2_CANAL	O74657 candida alb
5	1819	28.0	469	1	ALS4_CANAL	O74660 candida alb
6	940	14.5	1537	1	FLO1_YEAST	P32768 saccharomyc
7	932.5	14.4	1322	1	YAG3_YEAST	P39712 saccharomyc
8	721.5	11.1	5179	1	MUC2_HUMAN	Q02817 homo sapien
9	711	10.9	1075	1	FLO5_YEAST	P38894 saccharomyc
10	675.5	10.4	1367	1	AMYH_YEAST	P08640 saccharomyc
11	577	8.9	1161	1	DAN4_YEAST	P47179 saccharomyc
12	547.5	8.4	1609	1	FIG2_YEAST	P25653 saccharomyc
13	521.5	8.0	1140	1	YM96_YEAST	Q04893 saccharomyc
14	501	7.7	1169	1	YK82_YEAST	P36170 saccharomyc
15	464.5	7.2	670	1	VG50_HSVI1	Q00130 ictalurid h
16	446	6.9	1802	1	HKR1_YEAST	P41809 saccharomyc
17	443.5	6.8	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
18	418.5	6.4	725	1	AGA1_YEAST	P32323 saccharomyc
19	415.5	6.4	1306	1	MSB2_YEAST	P32334 saccharomyc
20	409.5	6.3	1041	1	EGT2_YEAST	P42835 saccharomyc
21	406	6.3	2090	1	HFC1_MESAU	P51611 mesocricetu
22	404.5	6.2	2035	1	HFC1_HUMAN	P51610 homo sapien
23	392	6.0	797	1	VGLX_HSVEB	P28968 equine herp
24	385.5	5.9	796	1	YS8A_CAEEL	Q09625 caenorhabdi
25	374.5	5.8	881	1	YJH8_YEAST	P47033 saccharomyc
26	353	5.4	650	1	SAG1_YEAST	P20840 saccharomyc
27	351.5	5.4	5376	1	ZAN_MOUSE	O88799 mus musculu
28	349.5	5.4	860	1	CHI2_COCIM	P54197 coccidioide
29	347.5	5.4	1150	1	APMU_PIG	P12021 sus scrofa
30	338	5.2	1848	1	CBPA_CLOCL	P38058 clostridium
31	323	5.0	907	1	VGP3_EBV	P03200 epstein-bar
32	320	4.9	886	1	VGP3_EBVA8	Q07284 epstein-bar
33	311	4.8	1253	1	DSPP_HUMAN	Q9nzw4 homo sapien
34	305	4.7	2090	1	N214_HUMAN	P35658 homo sapien
35	304.5	4.7	3396	1	PGCV_HUMAN	P13611 homo sapien
36	302	4.6	995	1	YIQ9_YEAST	P40442 saccharomyc
37	302	4.6	3562	1	PGCV_CHICK	Q90953 gallus gall
38	299.5	4.6	3381	1	PGCV_BOVIN	P81282 bos taurus
39	295.5	4.5	662	1	MUC1_XENLA	Q05049 xenopus lae
40	294	4.5	1849	1	IGA4_HAEIN	P45386 haemophilus
41	293.5	4.5	610	1	MUC4_HUMAN	Q99102 homo sapien
42	293	4.5	1199	1	N121_RAT	P52591 rattus norv
43	286	4.4	636	1	YNR6_YEAST	P53882 saccharomyc
44	286	4.4	937	1	HYR1_CANAL	P46591 candida alb
45	280.5	4.3	605	1	YHC8_YEAST	P38739 saccharomyc

ALIGNMENTS

```
RESULT   1
ALS1_CANAL
ID   ALS1_CANAL     STANDARD;      PRT;  1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST SAG1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L25902; AAC41649.2; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      486    491       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
```

```
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-THR.
FT   DOMAIN      874    877       POLY-SER.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;
```

Query Match 99.5%; Score 6465; DB 1; Length 1260;
Best Local Similarity 99.6%; Pred. No. 5.3e-272;
Matches 1255; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       301 SRSQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TDSIDTVVVQVPSPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Qy     481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy     541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Qy     601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Qy     661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720

Qy     721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy     781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840

Qy     841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     841 SSDGMLLSSTTLVTESETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900

Qy     901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960

Qy     961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020

Qy    1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080

Qy    1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140

Qy    1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 SVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200

Qy    1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
```

```
RESULT   2
ALS3_CANAL
ID   ALS3_CANAL     STANDARD;      PRT;  1119 AA.
AC   O74623;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 3 precursor.
GN   ALS3.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98309840; PubMed=9644209;
RA   Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;
RT   "Candida albicans ALS3 and insights into the nature of the ALS gene
RT   family.";
RL   Curr. Genet. 33:451-459(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U87956; AAC39486.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
```

```
FT   CHAIN        18   1119       AGGLUTININ-LIKE PROTEIN 3.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      557    563       POLY-THR.
FT   DOMAIN      593    597       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    777       POLY-THR.
FT   DOMAIN     1044   1047       POLY-THR.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    759    759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    845    845       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    987    987       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1050   1050       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1119 AA;  119927 MW;  6A3FB3FC8C879A71 CRC64;
```

```
  Query Match           64.7%;  Score 4204.5;  DB 1;  Length 1119;
  Best Local Similarity  64.9%;  Pred. No. 1.4e-174;
  Matches  844;  Conservative  106;  Mismatches  128;  Indels  223;  Gaps   18;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           ||||:||| :|||:|:|||||||:|||||||||| | :|||| ||||||||||||||
Db       1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPGTPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           |:|||||||||||||:|||||||||| |||||||| :|||| |||||||||:: |
Db      61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           |||| ||||||:|||||||||| ||||||||||||||||||| | |||:|:||:| |||
Db     121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
             ||  |||:||||||:|||||||| ||||||||||::: ||| |||||||:|||||||
Db     181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPPFIDAYISATDVNQYTLAYTNDYTCAG 300
           |||||||||||||||:|||| | |:||||||||:|||||||||| |||:| |:|||||
Db     241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
             | |||||||||:|||||||||||||||||||||||||||||:|: |||||||||:
Db     301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||||| ||||||||||||||:|||||||||:|||||:||| |||
Db     361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
           |||||||:|||| |||||:|||||||||||||:| ||| |||||:|||||||||||||||
Db     421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy     481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
```

```
           |:|: ||:| ||||||||||||:|||||||||||||||:|: ||:| ||||||||||||:|
Db     481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKE 540

Qy     541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
           ||| ||||||||||||| ||||||||||||||:|::|||||||||||||||||| ||||| |
Db     541 PPNHTVTTTEYWSQSYTTTTTVTAPPGGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIA 600

Qy     601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
           ||||||:|||||||| ||||||||||:||||||:| ||  ||||:|||||| ||||||||
Db     601 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW 660

Qy     661 SQSYATTTTITAPPGETDTVL----------------------------------IRE 684
           |||||||||||||||||||||                                   |||
Db     661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIRE 720

Qy     685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
           |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||: |
Db     721 PPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTIIA 780

Qy     745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSN 804
           ||| ||||:|||||||||||||||: | ||||||||||||||::|||||:||
Db     781 PPGETDTVLIREPPNPTVTTTEYWSQSYTTATTVTAPPGGTDTVIIYDTMSSSEIS---- 836

Qy     805 DITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELIC 864
                 ||||||| | ||                    | ||| | |::| || |
Db     837 -------SFSRPHYTNHTT----------------------LWSTTWVIETKTITETSC 866

Qy     865 SDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGV 924
              | || :| |: ||| |:: |: :||:|| : :| | |  |  ||        |:|||
Db     867 EGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTS-QSPSGI-------FSESGV 918

Qy     925 SVTTETS-VTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST 983
           || ||:| ||| ||                  |||  :||:| ||: |:|:::| ::| ||
Db     919 SVETESSTVTTAQTN----------------PSVPTTESEVVFTTKGNNENGPYESPST 961

Qy     984 STEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNG 1043
           :                | |: |  || |
Db     962 N----------------VKSSMDENSEFT---------------------------- 974

Qy    1044 DNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIE---ATT 1100
                      ||| ||                              |||:||     |
Db     975 ------------TSTAAS-----------------------------TSTDIENETIAT 992

Qy    1101 TSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNEN-GGKSPSTD 1159
           |   || || :||  | |          ||::||:: | :    | || | |||:||
Db     993 TGSVEASSPIISSSADET----------TTVTTTAESTSVI---EQPTNNNGGGKAPSA- 1038

Qy    1160 LTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQ 1219
                 |:  | :|:||:: | :|       : | :|| | |  |     |:| ||||||
Db    1039 -----TSSPSTTTTANNDSVITG--------TTSTNQSQSQSQYN-----SDTQQTTLSQ 1080

Qy    1220 QVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
           |:|||  | :  : :|:|||||:|||||| ||||||||||
Db    1081 QMTSSLVSLH--MLTTFDGSGSVIQHSTWLCGLITLLSLFI 1119
```

```
RESULT   3
ALA1_CANAL
ID   ALA1_CANAL     STANDARD;      PRT;  1419 AA.
AC   O13368;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein ALA1 precursor (Agglutinin-like adhesin).
GN   ALA1 OR ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98053977; PubMed=9393828;
RA   Gaur N.K., Klotz S.A.;
```

```
RT   "Expression, cloning, and characterization of a Candida albicans
RT   gene, ALA1, that confers adherence properties upon Saccharomyces
RT   cerevisiae for extracellular matrix proteins.";
RL   Infect. Immun. 65:5289-5294(1997).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF025429; AAB88883.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1419       AGGLUTININ-LIKE PROTEIN ALA1.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      437    441       POLY-THR.
FT   DOMAIN      673    676       POLY-SER.
FT   DOMAIN      687    690       POLY-SER.
FT   DOMAIN      700    703       POLY-SER.
FT   DOMAIN      719    724       POLY-SER.
FT   DOMAIN      749    752       POLY-SER.
FT   DOMAIN      787    791       POLY-SER.
FT   DOMAIN      869    872       POLY-SER.
FT   DOMAIN      875    883       POLY-SER.
FT   DOMAIN      901    911       POLY-SER.
FT   DOMAIN     1216   1221       POLY-SER.
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1301   1301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1326   1326       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1419 AA;  149635 MW;  249F33F688A9D5B6 CRC64;
```

```
  Query Match           49.2%;  Score 3194;  DB 1;  Length 1419;
  Best Local Similarity  49.5%;  Pred. No. 6.6e-131;
  Matches  708;  Conservative  182;  Mismatches  355;  Indels  186;  Gaps   29;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           |:|||||||||||| |:|| |||:|:| :|||||| ||||||||||||||||||||||||
Db       1 MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           ||||||| |||||||:| || ||||||||||||||||||||||:| |||| |:|
Db      61 ANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVNNNLRS 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           |||| |||||||||||||||| ||||||||||||||||||| | :|| | |||||||
Db     121 SIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGSKKLSIAVNFEKSTVDQ 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           | ||  || |||||: ||:||||||||||||||||:| |||||||||:||||:||:||
Db     181 SGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||:||:|||||||:|:| || | |||||||||||||||| :| ||| |:| |||||
Db     241 WNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVD 300

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
             | ||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 DYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
           |||||||||||| |||| :||::||:|| :|||:|    |||:||:||| | |||||:|
```

```
Db     421 TDSIDTVVVQVPSPNPTTTTTQFWSESFTSTTTITNSLKGTDSVIVREPHNPTVTTTEFW 480

Qy     481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
           |:||||| |:|: | ||||||:||| ||||||||:||:|:||| |:|  | ||||||:||
Db     481 SESFATTETITSKPEGTDSVIVREPHNPTVTTTEFWSESYATTETITNGPEGTDSVIVRE 540

Qy     541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
           | |||||||::||:|||| |:|  | ||||||::|| | ||||||:||:||||| |:|
Db     541 PHNPTVTTTKFWSESYATTETITNKPEGTDSVIVKEPYNPTVTTTEFWSESYATTETITN 600

Qy     601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTT--- 657
           | |||:||:||| | ||||||:||:|:||| |:|  | |||:::| :|   : :||
Db     601 GPEGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEESSSTTAIE 660

Qy     658 ----------EYWSQSYATTTTITAPPGETDTVLIREPPNHTVT---------------- 691
                     :  | |   :::|      :|  |  : |: : |
Db     661 SSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYDSD 720

Qy     692 ----------------------TTEYWSQS----YATTTTVTAPPGETDTVLIREPPNHT 725
                                || :|  |     :|: | ::      : |::   |
Db     721 SSSSIESSTLSSSDRCSSSISDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNSI 780

Qy     726 VTTTEYWSQSYATTTTVTAPPGGTDTVIIREP-----------PNPTVTTTEYW----- 768
           |: |  | |   ::|       || ::  :              |: |:::::  :
Db     781 STSQELSSSSSEESSTF-----ATDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIA 835

Qy     769 ------SQSFATTT--------TVTAPPGGT-DTVIIYESMSSSKISTSSNDI------- 806
                | || |:|       ::|: |  : |:     | |||  |:   :||
Db     836 GESDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFS 895

Qy     807 TSIIPSFS----------RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTES 856
           | ::|||           |||||||   |  ||||: :|:  |         |  ::||
Db     896 TLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESAN--DDTYTLSES 953

Qy     857 ETTTELICSDGKE---CSRLSSSSGIVTN-PDSNESSIVTSTVPTASTMSDSLSSTDG-I 911
             ||  | :|      | | :    |||  | |:  |  || | | |:   | : |
Db     954 TDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSAEQ 1013

Qy     912 SATSSDNV--SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTS 969
           | | : |:  |::  | :|::||:   |  | :  | ||  |||  | | |:: :: |:
Db    1014 SITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETS--TLSSDDSTS-SDTSISSTT 1070

Qy     970 NGD----NQSGTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSEPTNTR------ 1016
           | |    |   :| | ::  |  |  |    |:  |   :|| :|:| |   |
Db    1071 NSDTGNINAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELITTEL 1130

Qy    1017 -----EQPTTLSTTSNSITE------DITTSQPTGDNGDNTSSTNPVPTVAT----STLA 1061
                :| | ::| :| :|       : : |    ::   || |  | ||::    || |
Db    1131 TTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTEA 1190

Qy    1062 SA-----SEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTD 1116
           |      :   | : :  : :|||: :   ::  | |:    :  |    ||| |  ||
Db    1191 STATLGENSFSNVASTPSNIATSLRSTSSSSNHATESSGTVKSEASAEAIPSPPTS--TD 1248

Qy    1117 VTTEPTDTREQPTTLSTTSKTN-----SELVATTQATN---ENGGKSPSTDLTSSLTTGT 1168
                :  | : :  ||     |:: | | :|:   ||    : : |  ||
Db    1249 NRLSYSTEEAKGITYANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQ 1308

Qy    1169 SASTSANSELV----TSGSVTGGAVASASNDQSHSTSV---------------TNSNSI 1208
           : | :  | :    ::     | :|: |  |  |:|:                 | :|
Db    1309 PSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSA 1368

Qy    1209 VSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
           ||:  ||| ||| :|  | |  : :|| || |  :| :||   |::   |
Db    1369 ASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKHPSWLLKFISVALFF 1417
```

```
RESULT   4
ALS2_CANAL
ID   ALS2_CANAL     STANDARD;      PRT;   468 AA.
AC   O74657;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
```

```
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 2 precursor (Fragment).
GN   ALS2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF024580; AAC64235.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   >468       AGGLUTININ-LIKE PROTEIN 2.
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50203 MW;  BFE773E169ED0FAF CRC64;

  Query Match             30.7%;  Score 1992.5;  DB 1;  Length 468;
  Best Local Similarity   80.2%;  Pred. No. 1.2e-79;
  Matches  376;  Conservative   33;  Mismatches   59;  Indels    1;  Gaps    1;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           || || || | :|:|:|| |||||:||:||||: | |||:||| ||||||||||||||
Db       1 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           ||||||||||||||: | ||||||||:|||||||||||||||||:| |||:: | |
Db      61 ANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 120

       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           |||| |||||||:|||||||| ||| |:|| |||||||||||| |||||:|||:  |
Db     121 SIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           | |  |||::||:|||:  :||||| |||||| ||| |    || |||||:|:|||||||
Db     181 SGYFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||:||||:| || |||:| || | |:|||||||| | | | :  |:  | |||| | |
Db     241 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVG 299

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
           | |||||| ||  || ||:| ||| ||||||||||||||||||||||||||||||||||||
Db     300 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 359

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     360 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 419

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
           |||||||||||| |||||:||||||||:||||||||||||||:|||||
Db     420 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREP 468
```

```
RESULT   5
ALS4_CANAL
ID   ALS4_CANAL     STANDARD;      PRT;   469 AA.
AC   O74660;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 4 precursor (Fragment).
GN   ALS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF024584; AAC64239.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   >469       AGGLUTININ-LIKE PROTEIN 4.
FT   NON_TER     469    469
SQ   SEQUENCE   469 AA;  49604 MW;  0EDCAB19B89EFCB1 CRC64;

  Query Match             28.0%;  Score 1819;  DB 1;  Length 469;
  Best Local Similarity   72.1%;  Pred. No. 3.7e-72;
  Matches  338;  Conservative   48;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           || || || | :|:|:|| ||||:|||||||:|||:| ::|| ||| ||:|||||| :
Db       1 MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPATPTWTAVIGWSLDGAT 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           |: |||||:|||||: | |||:|| |||  ||||   | |||||||:::|||   : :
Db      61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
             || |||||| :|:|||:||  || :|:||||| |||||||||  ||  |:||||||
Db     121 DTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVAS 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           |  :  ||::|||::   ||:  :| ||||||||||::     |||| :|:||: ||||
Db     181 SDRILLSRILPSLSQAVNLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||||:||||||||||||  : : :|||||||||:|||||| |: ||: ||| | | |
Db     241 WNYPISSESFSYTKTCTSTSVLVTFQNVPAGYRPFVDAYISATRVSSYTMQYTNIYACVG 300

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
           :      ||  | || || |||||| || ||||||||||||||||||   ||||||||||
Db     301 AASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSDTDKTKTIEILQ 360

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
```

```
          ||||||||||| |||||:||||||||:|||||||||||||||:||||||
Db    421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREP 469


RESULT   6
FLO1_YEAST
ID   FLO1_YEAST     STANDARD;      PRT;  1537 AA.
AC   P32768;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Flocculation protein FLO1 precursor (Flocculin 1).
GN   FLO1 OR YAR050W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94262325; PubMed=8203162;
RA   Watari J., Takata Y., Ogawa M., Sahara H., Koshino S., Onnela M.-L.,
RA   Airaksinen U., Jaatinen R., Penttilae M., Keraenen S.;
RT   "Molecular cloning and analysis of the yeast flocculation gene FLO1.";
RL   Yeast 10:211-225(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN   [3]
RP   PRELIMINARY SEQUENCE FROM N.A.
RX   MEDLINE=93289821; PubMed=8511970;
RA   Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,
RA   Steensma H.Y.;
RT   "Sequence of the open reading frame of the FLO1 gene from
RT   Saccharomyces cerevisiae.";
RL   Yeast 9:423-427(1993).
RN   [4]
RP   REVIEW.
RX   MEDLINE=96076625; PubMed=7502576;
RA   Teunissen A.W., Steensma H.Y.;
RT   "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT   constitute a new subtelomeric gene family.";
     Yeast 11:1001-1013(1995).
     -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
     -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78160; CAA55024.1; -.
DR   EMBL; L28920; AAC09499.1; ALT_SEQ.
DR   PIR; S31230; S31230.
DR   SGD; S0000084; FLO1.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 18.
KW   Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       FLOCCULATION PROTEIN FLO1.
FT   PROPEP        ?   1537       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      274   1240       SER/THR-RICH.
FT   DOMAIN      278   1087       18 X 45 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      278    322       1-1.
FT   REPEAT      323    367       1-2.
FT   REPEAT      368    412       1-3.
FT   REPEAT      413    457       1-4.
FT   REPEAT      458    502       1-5.
FT   REPEAT      503    547       1-6.
FT   REPEAT      548    592       1-7.
FT   REPEAT      593    637       1-8.
FT   REPEAT      638    682       1-9.
FT   REPEAT      683    727       1-10.
FT   REPEAT      728    772       1-11.
FT   REPEAT      773    817       1-12.
FT   REPEAT      818    862       1-13.
FT   REPEAT      863    907       1-14.
FT   REPEAT      908    952       1-15.
FT   REPEAT      953    997       1-16.
FT   REPEAT      998   1042       1-17.
FT   REPEAT     1043   1087       1-18.
FT   DOMAIN     1118   1157       2 X 20 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1118   1137       2-1.
FT   REPEAT     1138   1157       2-2.
FT   DOMAIN     1226   1392       3 X 51 AA APPROXIMATE REPEATS.
FT   REPEAT     1226   1276       3-1.
FT   REPEAT     1291   1341       3-2.
FT   REPEAT     1342   1392       3-3.
FT   DOMAIN     1408   1434       3 X 9 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1408   1416       4-1.
FT   REPEAT     1417   1425       4-2.
FT   REPEAT     1426   1434       4-3.
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    374    374       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    509    509       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    644    644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    734    734       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1537 AA;  160722 MW;  992650C6BE9A8CEA CRC64;


  Query Match           14.5%;  Score 940;  DB 1;  Length 1537;
  Best Local Similarity   26.0%;  Pred. No. 1.1e-33;
  Matches  417;  Conservative  244;  Mismatches  523;  Indels  418;  Gaps   85;

Qy      2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
          |  |||| |  |:|| |           :|:' :|      :| |:|||| |:
Db     10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy     41 ----------------------------KGPGYPTW-----------NAVLGWSLD--G 58
                                       :  | |            :  || |  |
Db     69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy     59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
              | :  || |            ||: |                          || :
Db    129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy     86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
            | ||:|           |   |:|  :  |          : |:  ::||    ||||
Db    188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy    132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
            ||  | | |              :|:|      |:     || ||| |   |
Db    239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy    191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
          :    | | :   |    :|| | :    : |        |      ||   :| |
```

```
Db    278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy    251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310

Db    338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy    311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT---------- 355

Db    385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGL 438

Qy    356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398

Db    439 PTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTEMTTVTGTNGLPTDET-IIVIRTPTT 497

Qy    399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447

Db    498 ATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 556

Qy    448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494

Db    557 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTELTTVTGTN 616

Qy    495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542

Db    617 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 676

Qy    543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591

Db    677 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDT 736

Qy    592 YATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTVT--- 635

Db    737 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTN 796

Qy    636 GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686

Db    797 GQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVIRTPT 856

Qy    687 NHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEYWSQS 735

Db    857 SEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGT 916

Qy    736 YATT----TTVTAPPG-GTD--TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPPGGT- 785

Db    917 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTITGTN 976

Qy    786 -----DTVIIYESMSSSKI------------STSSNDITSIIPSFSRP------------ 816

Db    977 GQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPT 1036

Qy    817 --HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICSDGKE 869

Db   1037 SEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQI 1096

Qy    870 CSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927

Db   1097 TSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTSTSIF 1155

Qy    928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987

Db   1156 SESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS----L 1205

Qy    988 EIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITEDITT 1035

Db   1206 PLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISPAIVS 1256

Qy   1036 SQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093

Db   1257 TATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCSKTAS 1316

Qy   1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELVATT- 1145

Db   1317 PAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAIVSTA 1374

Qy   1146 ----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSVTGGA 1188

Db   1375 TATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITNTGAA 1434

Qy   1189 -------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224

Db   1435 ETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1494

Qy   1225 SP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260

Db   1495 PASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536


RESULT   7
YAG3_YEAST
ID   YAG3_YEAST     STANDARD;      PRT;  1322 AA.
AC   P39712;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Hypothetical 138.1 kDa protein in FLO9-GDH3 intergenic precursor.
GN   YAL063C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC   -!- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12980; AAC04971.1; -.
DR   SGD; S0000059; YAL063C.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 13.
KW   Hypothetical protein; Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1322       HYPOTHETICAL PROTEIN YAL063C.
FT   TRANSMEM    366    388       POTENTIAL.
FT   TRANSMEM    754    775       POTENTIAL.
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    257    257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    270    270       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    464    464       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    509    509       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    644    644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    734    734       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    888    888       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1322 AA;  138072 MW;  AADFD1FF13267CEA CRC64;
```

```
Query Match           14.4%;  Score 932.5;  DB 1;  Length 1322;
Best Local Similarity  28.1%;  Pred. No. 2e-33;
Matches  400;  Conservative  215;  Mismatches  468;  Indels  339;  Gaps   79;

Qy      28 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQT----S 83
           :| |:|||| |:   || :      |        :   |:|||   ::| |
Db      50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCV---SSSGTFPCPQ 98

Qy      84 VDLTAD-GVK-YATCQ------FYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIA-- 133
            ||  : | |    |       ::| : |  ::| |              |||  :
Db      99 EDLYGNWGCKGIGACSNNPIIAYWSTDLFGFYTTPT---------------NVTLEMTGY 143

Qy     134 FNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSL 193
           |    ||| |                          |: :||| || |
Db     144 FLPPQTGSYT----------------------------FKFATVDDSAIL--------- 165

Qy     194 NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC----------SNIHIGITKGLNDWN- 242
                                  |  | :| :|          :|  |    |:  ||
Db     166 ----------------------SVGGSIAFECCAQEQPPITSTNFTI---NGIKPWNG 198

Qy     243 -YPVSSESFSYTKTCTSNGIQIKYQNVPA-GYRPF-------------IDAYISATDVN- 286
            | :         |         ::| | |  | |  |             : |:   | |
Db     199 SPPDNITGTVYMYAGFYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEGYVYTFDNNL 258

Qy     287 ---QYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDA-------GSNGI-----VIVATT 331
               |:   ::|| : :   ::|    |||   | :        |:|||:     |||  |
Db     259 SQPNCTIPDPSNYTVSTTITTTEP----WTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy     332 RTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIT-TSYVGVTTSYLTKTAPIGETAT 390
           | | ||   || |:| :   | |          ||:| |: |    : :    |   |
Db     315 PT-TASTIITTTEPWNSTFTSTST-------ELTTVTGTNGVRTDETIIVIRTPTTATTA 366

Qy     391 VIVDVPYHTTTTVTSEWTGTITTTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
           :   |:::| | ||    |  ||  | ||  |||    :|:: |    | ::||: |: :
Db     367 ITTTEPWNSTFTSTS----TELTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 421

Qy     448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
           | :|    ||||    |  ||   ::|| |   |  :|||: |: :| :|    ||||
Db     422 FTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTN 481

Qy     495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
           |  ||   ::|| |   |  :|||: |: :| :|    ||||   |  ||   ::|| |
Db     482 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPT 541

Qy     543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
           |  :|||: |: :: :|    ||||   |  ||   ::|| |   |  :|||| |: :
       542 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNST 601

       592 YATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTGP- 637
           : :|    ||||    |  ||   ::|| |   |  :|||: |: :| :|    |||||
Db     602 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTN 661

Qy     638 --PSGTDTVIIREPPNPT--VTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
             |:    ::|| |   |  :|||: |: :: :|    ||:|     | :   ::|| |
Db     662 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPT 721

Qy     687 NHT--VTTTEYWSQSYATT----TTVTAP---PGETDTVLIREPPNHTV--TTTEYWSQS 735
           |  :|||: |: :: :|    ||||     | :   ::|| | :   :  ||||  |: :
Db     722 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGT 781

Qy     736 YATT----TTVTAPPG-GTD--TVIIREPPNP--TVTTTEYWSQSFATT----TTVTAPP 782
           : :|    ||||    |  ||   ::|| | :     |||| |: :| :|    ||:|
Db     782 FTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTN 841

Qy     783 G--GTDTVIIYESMS---SSKISTSSNDITSIIPS---FSRPHYVNSTTSDLSTFESSSM 834
           |    :|||| :: :   || :|:|| |||  | |      | | :: || :|:   ||
Db     842 GQPTDETVIIVKTPTTAISSSLSSSSGQITSFITSARPIITPFYPSNGTSVISSSVISSS 901

Qy     835 NTPTSISSDGM---LLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIV 891
           :| : : |   :   |::|: ::: |  :: :| |      | || ||
Db     902 DTSSLVISSSVTSSLVTSSPVISSSFISSPVISSTTTSASILSESS----------K 948
```

```
Qy     892 TSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQ 951
           :| :||:|: | |  |  | :::::|   | | :|  :  |  :  : | |::|: ||
Db     949 SSVIPTSSSTSGSSESETGSASSAS---SSSSISSESPKSTYSSSSLP-PVTSATTSQEI 1004

Qy     952 LSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE 1011
           ||:| |: :::             | | :  ||:  : |    :||    |:
Db    1005 TSSLPPVTTTKT-----------------SEQTTLVTVTSCESHVCTESISSAIVSTAT 1046

Qy    1012 PT---NTREQPTTLSTTSNSITEDIT-TSQPTGDNGDNTSSTNPVPTVATSTLASASEED 1067
           |    | | |     ::  ||:  | |:: |    : |: |    || |    |: | |
Db    1047 VTVSGATTEYTTWCPISTTEITKQTTETTKQTKGTTEQTTETTKQTTVVT---ISSCESD 1103

Qy    1068 --NKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTR 1125
             :|: |    |||     |  :  ||   |  | :          | |: | :| |
Db    1104 VCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESKQQTTLVTVTSCGSGVCSETT--- 1160

Qy    1126 EQPTTLSTTSKTNSELVAT-----TQATNENGGKSPSTDLTSSLTTGTSAST------SA 1174
             |  :|| : | :::|        | |||      |     ||  || | |:          :|
Db    1161 -SPAIVSTATATVNDVVTVYSTWRPQTTNEQSVSSKMNSATSETTTNTGAAETTTSTGAA 1219

Qy    1175 NSELVTSGSVT-----GGAVASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
           :: | : |::          |||::   ||:||     | :   ::::  :|:|||  |:
Db    1220 ETKTVVTSSISRFNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1279

Qy    1225 ------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
                 || :   ||| || : :   : |   |  | | |
Db    1280 PASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1321
```

```
RESULT   8
MUC2_HUMAN
ID   MUC2_HUMAN     STANDARD;      PRT;  5179 AA.
AC   Q02817; Q14878;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 2 precursor (Intestinal mucin 2).
GN   MUC2 OR SMUC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=94132002; PubMed=8300571;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT   "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT   Identification of the amino terminus and overall sequence similarity
RT   to prepro-von Willebrand factor.";
RL   J. Biol. Chem. 269:2440-2446(1994).
RN   [2]
RP   SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=93016075; PubMed=1400449;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA   Kim Y.S.;
RT   "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT   both upstream and downstream of its central repetitive region.";
RL   J. Biol. Chem. 267:21375-21383(1992).
RN   [3]
RP   SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX   MEDLINE=91358717; PubMed=1885763;
RA   Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA   Petersen G.M., Kim Y.S.;
RT   "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT   and polymorphism.";
RL   J. Clin. Invest. 88:1005-1013(1991).
CC   -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC       OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC       PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC       AGENTS AT MUCOSAL SURFACES.
```

```
CC   -!- SUBUNIT: MULTIMERIC.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC       BRONCHUS, CERVIX AND GALL BLADDER.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC       INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES.
CC   -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC       OF SILKWORM HEMOCYTIN.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L21998; AAB95295.1; -.
DR   EMBL; M74027; AAA59875.1; -.
DR   EMBL; M94131; AAA59163.1; -.
DR   EMBL; M94132; AAA59164.1; -.
DR   MIM; 158370; -.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00094; vwd; 4.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   SMART; SM00214; VWC; 2.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   5179       MUCIN 2.
FT   DOMAIN    1401   1747       APPROXIMATE REPEATS.
FT   REPEAT    1401   1416       1.
FT   REPEAT    1417   1432       2.
FT   REPEAT    1433   1448       3.
FT   REPEAT    1449   1464       4.
FT   REPEAT    1465   1471       5.
FT   REPEAT    1472   1478       6.
FT   REPEAT    1479   1494       7A.
FT   REPEAT    1495   1517       7B.
FT   REPEAT    1518   1533       8A.
FT   REPEAT    1534   1556       8B.
FT   REPEAT    1557   1572       9A.
FT   REPEAT    1573   1596       9B.
FT   REPEAT    1597   1612       10A.
FT   REPEAT    1613   1635       10B.
FT   REPEAT    1636   1651       11A.
FT   REPEAT    1652   1675       11B.
FT   REPEAT    1676   1683       12.
FT   REPEAT    1684   1699       13.
FT   REPEAT    1700   1715       14.
FT   REPEAT    1716   1731       15.
FT   REPEAT    1732   1747       16.
FT   DOMAIN    4815   4886       VWFC 1.
FT   DOMAIN    4924   4991       VWFC 2.
FT   DOMAIN    5075   5160       CTCK.
FT   DISULFID  5075   5122       BY SIMILARITY.
FT   DISULFID  5089   5136       BY SIMILARITY.
FT   DISULFID  5098   5152       BY SIMILARITY.
FT   DISULFID  5102   5154       BY SIMILARITY.
FT   DISULFID     ?   5159       BY SIMILARITY.
FT   CARBOHYD   163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   670    670       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   770    770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   894    894       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1139   1139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1154   1154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1215   1215       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1230   1230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1246   1246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1787   1787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1820   1820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4339   4339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4351   4351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4362   4362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4373   4373       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4422   4422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4438   4438       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4502   4502       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4616   4616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4627   4627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4752   4752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4787   4787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4881   4881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4888   4888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4955   4955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4970   4970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  5019   5019       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  5038   5038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  5069   5069       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT  1351   1351       H -> L (IN REF. 3).
FT   CONFLICT  1412   1412       T -> S (IN REF. 3).
FT   CONFLICT  1449   1449       L -> P (IN REF. 3).
FT   CONFLICT  1504   1504       M -> T (IN REF. 3).
FT   CONFLICT  4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE  5179 AA;  540295 MW;  85CD7571FB9A5663 CRC64;
```

```
  Query Match           11.1%;  Score 721.5;  DB 1;  Length 5179;
  Best Local Similarity  25.3%;  Pred. No. 1.1e-23;
  Matches  335;  Conservative  158;  Mismatches  554;  Indels  279;  Gaps   54;

Qy      59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118
           |:  |  | | : |     ||:  |  :|       |         |  | :| | :
Db    1647 TTTTPPPTTTPSPP-----TTTTPSPPITTTTTPPPTT--------TPSSPITTTPSPPT 1693

Qy     119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
           :      | |         | |||        |  | : |         ::      :|:
Db    1694 TTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTM------TTL 1747

Qy     179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224
           |   |: |  : : ||:   |          |    | |   |:  ||   |    ||
Db    1748 PPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTE 1807

Qy     225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272
           :     ||   |  |  :|  :  |          | |  | |:  | ::  | |
Db    1808 L------IGDVCG-PGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGV 1860

Qy     273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332
           |             : | |  :  |      ::| |:- |  :|    :   :   ||
Db    1861 IPM----------AFCLNYEINVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTV 1907

Qy     333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIG-ETATV 391
           | | : |   |    |    |       |  | | ||: :  ||:      | | :| |
Db    1908 TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT 1967

Qy     392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432
           |  ||||||   | |          |||||          |:|  |  | |     |
Db    1968 ---TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP 2024

Qy     433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492
           | || : |   :    ||||||  |  | |   : |     :|||         || | |
Db    2025 TPTPTGTQTP-TTTPITTTTTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTP 2072
```

```
Qy    493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535
Db   2073 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT 2132

Qy    536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592
Db   2133 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-----PTTTPITTT------- 2180

Qy    593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643
Db   2181 -TTVTPTPTPTGTQT-----PTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTT 2226

Qy    644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703
Db   2227 TTVTPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT--------QTPTTT-----PITTT 2272

Qy    704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754
Db   2273 TTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTTTTV 2321

Qy    755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814
Db   2322 TPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT----QTPTTTPITTT----TTVTPTPT 2372

Qy    815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874
Db   2373 PTGTQTPTTTPITT---TTTVTPTP-TPTGTQTPTTTPITTTTTVTPTPTPTGTQ----- 2423

Qy    875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932
Db   2424 -------TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2476

Qy    933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980
Db   2477 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2536

Qy    981 QSTSTEIEIVTTSSTKVLP-----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028
Db   2537 TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT 2596

Qy   1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087
Db   2597 VT---PTPTPT---GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT 2650

Qy   1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ----------- 1127
Db   2651 GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT 2710

Qy   1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
Db   2711 TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP 2770

Qy   1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
Db   2771 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2830

Qy   1236 YDGSGS 1241
Db   2831 PTPTGT 2836
```

RESULT 9
FLO5_YEAST
ID FLO5_YEAST STANDARD; PRT; 1075 AA.
AC P38894;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Flocculation protein FLO5 precursor (Flocculin 5).
GN FLO5 OR YHR211W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
RN [2]
RP REVIEW.
RX MEDLINE=96076625; PubMed=7502576;
RA Teunissen A.W., Steensma H.Y.;
RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT constitute a new subtelomeric gene family.";
RL Yeast 11:1001-1013(1995).
CC -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U00029; AAB69731.1; -.
DR PIR; S48992; S48992.
DR SGD; S0001254; FLO5.
DR InterPro; IPR001389; Flocculin.
DR Pfam; PF00624; Flocculin; 8.
KW Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 ? FLOCCULATION PROTEIN FLO5.
FT PROPEP ? 1075 REMOVED IN MATURE FORM (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 663 663 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 749 749 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1075 AA; 111981 MW; D151B370B60C8D9F CRC64;

Query Match 10.9%; Score 711; DB 1; Length 1075;
Best Local Similarity 25.9%; Pred. No. 5.5e-24;
Matches 324; Conservative 192; Mismatches 393; Indels 342; Gaps 66;

```
Qy    153 AGTNTVTFNDGDKDISIDVEF------EKSTVDPSAYL---YASRV-------------- 189
Db     23 SGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKTKLGSVGGQTDISID 82

Qy    190 --MPSLNKVTTLFVAPQCENGYTS-GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
Db     83 YNIPCVSSSGT-FPCPQ-EDSYGNWGCKGMGA--------CSN-----SQGIAYW----S 123

Qy    247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR---L 303
Db    124 TDLFGFYTTPTNVTLEMTGYFLPPQTGSYTFSFATVDDSAILSVGGSIAFECCAQEQPPI 183

Qy    304 QSKPFTLRWTGYKNSDAGSNGIVIVATT----------RTVTDSTTAVTTLPFNPSVDKT 353
```

```
Db    184 TSTNFTI--NGIKPWD-GSLPDNITGTVYMYAGYYYPLKVVYSNAVSWGTLPI--SVE-- 236

Qy    354 KTIEILQPIPT-TTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT---TTVTSEWTG 409

Db    237 --------LPDGTTVSDNFEGYVYSF---DDDLSQSNCTIPDPSIHTTSTITTTTEPWTG 285

Qy    410 TIT------TTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDT 463

Db    286 TFTSTSTEMTTITDTNGQLTDETVIV---IRTPT------------TASTIT-------- 322

Qy    464 VIIREPPNHTVTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTE 514

Db    323 ----------TTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTE 371

Qy    515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTEYWSQSYATTTTVTAP 565

Db    372 PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEVTT 431

Qy    566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNH--TVTTTEY 623

Db    432 ITGTNG----QPTDETV------------------------IVIRTPTSEGLITTTTEP 462

Qy    624 WSQSFATT----TTVT---GPPSGTDTVIIREPPNPTV--TTTEYWSQSYATT----TTI 670

Db    463 WTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 522

Qy    671 TAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPGE-TD--TVLI 718

Db    523 TGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 582

Qy    719 REPPNH--TVTTTEYWSQSYATTTTVTAPPGGT------DTVIIREPPNPTVTTTEYWSQ 770

Db    583 RTPTSEGLITRTTEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPT-TAISSSLSSS 641

Qy    771 SFATTTTVTA---------PPGGTDTV---IIYESMSSSKISTSSNDITSIIPSFSRPHY 818

Db    642 SGQITSSITSSRPIITPFYPSNGTSVISSSVISSSVTSSLVTSSSFISSSVISS------ 695

Qy    819 VNSTTSDLSTFESSSMNT--PTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSS 876

Db    696 --STTTSTSIFSESSTSSVIPTSSSTSGSSESKTSSASSSSSSSSISSESPKSPTNSSSS 753

Qy    877 SGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS------SDNVSKSGVSVTTET 930

Db    754 LPPVTSATTGQET--ASSLPPATTTKTS-EQTTLVTVTSCESHVCTESISSAIVSTATVT 810

Qy    931 --SVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE 988

Db    811 VSGVTTEYTTWCPISTTETT----------------------KQTKGTTEQTKGTTEQT 847

Qy    989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048

Db    848 TETTKQTTVV-TISSCESDICSKTAS----PAIVSTSTATI------------NGVTTEY 890

Qy   1049 TNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP- 1107

Db    891 TTWCPISTTES----------------------KQQTTLVTVTSCESGVCSETTSPA 925

Qy   1108 --SPAVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL 1160

Db    926 IVSTATATVNDVVTVYPTWRPQTTNEQ----SVSSKMNS---ATSETTTNTGAAETKTAV 978

Qy   1161 TSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSV-----TNSNSIVSNTPQT 1215

Db    979 TSSLSRFNHAETQ--------------TASATDVIGHSSSVVSVSETGNTMSLTSSGLS 1023

Qy   1216 TLSQQVTSS------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260

Db   1024 TMSQQPRSTPASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1074


RESULT 10
AMYH_YEAST
```

```
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;


  Query Match             10.4%;  Score 675.5;  DB 1;  Length 1367;
  Best Local Similarity   26.2%;  Pred. No. 2.5e-22;
  Matches  378;  Conservative  177;  Mismatches  535;  Indels  353;  Gaps   65;
```

```
Qy      1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db      1 MQRPFLLAYLVLSLL-----------FNSAL-----------GFPTALVPRG-SSEGTS 36

Qy     61 AN------PGDTFTLNM--PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTC 112
Db     37 CNSIVNGCPNLDFNWHMDQQNIMQYTLDVTSVSWVQDNT-YQITIHVKGKE--------- 86

Qy    113 TVNDALKS--SIKAFGTVTLPIAFNVGGTGSSTDLEDSKC-FTAGTNTVTFN-DGDKDI- 167
Db     87 --NIDLKYLWSLKIIGVTGPKGTVQLYGYNENTYLIDNPTDFTA-----TFEVYATQDVN 139

Qy    168 -------SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 220
Db    140 SCQVWMPNFQIQFE--------YLQGS--------------AAQYASSWQWGTTSFDLST 177

Qy    221 GDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI 280
Db    178 G-----CNNYD---NQGHSQTDFP----GFYWNIDCDNN--------------------C 205

Qy    281 SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA 340
Db    206 GGTKSSTTTSSTSESSTTTSSTSESSTTT-------SSTSESS----TTTSSTSESSTSS 254

Qy    341 VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT 400
Db    255 STTAPATPT-----TTSCTKEKPTPPTTTS--------CTKEKP---------TPPHHDT 292

Qy    401 TTVTSEWTGTITTTTTRTN---PTDSIDTV-VVQVPLPNPTVSTTEYWSQ--SFATTTTV 454
Db    293 TPCTKKKTTTSKTCTKKTTTPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESS 352

Qy    455 TAP---PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVT 511
Db    353 SAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAP-----------VPTPSSS 401

Qy    512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTT-VTAPPGGTD 570
Db    402 TTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTP 461

Qy    571 SVIIREPPNHTVT--TTEYWS-----QSYATTTTVTAPPGGTDTVIIREP-PNHTVTTTE 622
Db    462 SSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE 521

Qy    623 YWS-----QSFATTTTVTGPPSGTDTVIIREP-PNPTVTTTEYWSQSYATTTTITAPPGE 676
Db    522 SSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE------SSSTPVTSSTTE 575

Qy    677 TDTVLIREPPNHTVTTTEYWSQSYAT----TTTVTAPPGETDTVLIREPPNHTVT--TTE 730
Db    576 SSSAPV---PTPSSSTTESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTE 632

Qy    731 YWS-----QSYATTTTVTAP---PGGTDTVIIREP-PNPTVTTTEYWSQSFATTTTVTAP 781
Db    633 SSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE--SSSAPVTSSTTES 690

Qy    782 PGGTDTVIIYESMSS-----SKISTSSNDITSIIPSFSR--------PHYVNSTTSD--- 825
Db    691 SSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSA 750

Qy    826 ---LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
Db    751 PVTSSTTESSSAPVPTP-SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPV 809

Qy    883 PDSNESSIVTSTVPTASTMSDSL-SSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNP 941
Db    810 PTPSSSSNITSSAPSSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAPVPTP 869

Qy    942 LSSSVTSLTQLSSIPSVSESES-----KVTFTSNGDNQSGTHDSQSTSTEIEIV------ 990
Db    870 SSSSNITSSAPSSIPFSSTTESFSTGTTVTPSSSKYPGSQTETSVSSTTETTIVPTKTTT 929

Qy    991 ---TTSSTKVLPPVVSSNTDLTSEPTN-------TREQPTTLST--TSNSITEDITTSQP 1038
```

```
Db    930 SVTTPSTTTITTTVCSTGTNSAGETTSGCSPKTVTTTVPTTTTSVTTSSTTTITTTVCS 989

Qy   1039 TGDN--GDNTSSTNPVPTVATSTLASASEEDNKSGSHESA-----------------ST 1078
Db    990 TGTNSAGETTSGCSP-KTITTTVPCSTSPSETASESTTTSPTTPVTTVVSTTVVTTEYST 1048

Qy   1079 SLKPSMGENSGLTTS--TEIEATTTSPTEAPSPAV----------------SSGTD---- 1116
Db   1049 STKPG-GE---ITTTFVTKNIPTTYLTTIAPTPSVTTVTNFTPTTITTTVCSTGTNSAGE 1104

Qy   1117 ----------VTTEPTDT-----REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLT 1161
Db   1105 TTSGCSPKTVTTTVPCSTGTGEYTTEATTLVTTAVTTT-VVTTESSTGTNSAGKTTTGYT 1163

Qy   1162 SSLTTGTSASTSANSELVTSG------SVTGGAVASASNDQSHSTSVTNSNSIVSN---- 1211
Db   1164 TKSVPTTYVTTLAPSAPVTPATNAVPTTITTTECSAATNAAGETTSVCSAKTIVSSASAG 1223

Qy   1212 --------TPQTT---LSQQVTSSSPSTN----------TFIASTYDGSGSIIQ 1244
Db   1224 ENTAPSATTPVTTAIPTTVITTESSVGTNSAGETTTGYTTKSIPTTYITTLIPGSNGAKN 1283

Qy   1245 HST 1247
Db   1284 YET 1286
```

RESULT 11
DAN4_YEAST
ID DAN4_YEAST STANDARD; PRT; 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
GN DAN4 OR YJR151C OR J2223.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Scarcez T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP REGULATION.
RX MEDLINE=21113168; PubMed=11160904;
RA Cohen B.D., Sertil O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites.";
RL Nucleic Acids Res. 29:799-808(2001).
CC -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SRP1 / TIP1 FAMILY.
CC ----------

CC ----------
DR EMBL; Z49651; CAA89684.1; -.
DR SGD; S0003912; DAN4.
DR InterPro; IPR000992; SRP1_TIP1.
DR Pfam; PF00660; SRP1_TIP1; 1.
DR PROSITE; PS00724; SRP1_TIP1; 1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.

```
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1146       CELL WALL PROTEIN DAN4.
FT   PROPEP     1147   1161       REMOVED IN MATURE FORM (POTENTIAL).
FT   LIPID      1146   1146       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   1161 AA;  118358 MW;  7954C15D69F0CA58 CRC64;

  Query Match             8.9%;  Score 577;  DB 1;  Length 1161;
  Best Local Similarity   26.3%;  Pred. No. 3.6e-18;
  Matches  293;  Conservative  131;  Mismatches  427;  Indels  264;  Gaps   46;

Qy    276 IDAYISATDVNQYTLAY--------TNDY--TCAGSRLQSKPFTLRWTGYKNSDAGSNGI 325
          |:  :  :|:  :     |          |   |     | :        || | ||     :
Db     36 IELAVYVSDIRAHIFQYYSFRNHHKTETYPSEIAAAVFDYGDFTTRLTGISGDE------ 89

Qy    326 VIVATTRTVTDSTTAVTTLPFNPSVDKTKTIE-ILQPIPTTITTSYVGVTTSYLTKTAP 384
                  || :|       | |   |::    : : |   |||:| ||      ||   | | |
Db     90 ----VTRMITGVPWYSTRL--KPAISSALSKDGIYTAIPTSTSTT-----TTKSSTSTTP 138

Qy    385 IGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYW 444
                         |||:||      ||:|| | || |    :             || |||
Db    139 ---------------TTTITS------TTSTTSTTPTTSTTST-------TPTTSTT--- 167

Qy    445 SQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 504
          | :   |:|| |  |     | :               |  ||    | :  |:|| | |    | |
Db    168 STTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTT 227

Qy    505 PPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTA 564
               || :||    ||:    :|| |      |          || :||          |:|| ||
Db    228 STTPTTSTTSTTSQTSTKSTTPTTSSTST---------TPTTSTTP------TTSTTSTA 272

Qy    565 PPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 624
           |    | |            | :||    : | | ||: |:       | :          :  ::||
Db    273 PTTSTTS---------TTSTTS--TISTAPTTSTTSSTFSTSSA----SASSVISTTATT 317

Qy    625 SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684
           | :||: ||           :|:||    :: |   ||||                :|| :
Db    318 STTFASLTTPATSTASTDHT------TSSVSTTNAFTTSATTTTT-------SDTYISSS 364

Qy    685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQ---------- 734
           |:    :: |   :  |   |::        |      :           ||        | :|: |
Db    365 SPSQVTSSAEPTTVSEVTSS--VEPTRSSQVTSSAEP----TTVSEFTSSVEPTRSSQVT 418

Qy    735 SYATTTTV---TAPPGGTDTVIIREPPNPTVTTTEYWSQ----------SFATTTTVTAP 781
           | |    |||      |:         :   :         || | :|: |       | |   |||:
Db    419 SSAEPTTVSEFTSSVEPTRSSQVTSSAEPT-TVSEFTSSVEPTRSSQVTSSAEPTTVSEF 477

Qy    782 PGGTDTVIIYESMSSSKISTSSNDITSIIP-------SFSRPHYVNSTTSDLSTFESSSM 834
               :        :    | |:: :| |   :|: |          | : |   |:   || :     || :
Db    478 TSSVEPTRSSQVTSSAEPTTVSEFTSSVEPIRSSQVTSSAEPTTVSEVTSSVEPIRSSQV 537

Qy    835 NTPTSISSDGMLLSSTTLVTE----SETTT-------------ELICSDGKECSRLSSS 876
            |    :||   |     |      |      |: ||                  ||| |      | ::||
Db    538 TTTEPVSSFGSTFSEITSSAEPLSFSKATTSAESISSNQITISSELIVS-----SVITSS 592

Qy    877 SGIVTNPDSNESSIVTSTV-PTA---STMSDSLSSTDGISATS---------------- 915
          | |  :: :     || ::|:| ||:       :   :|:|||: :||||
Db    593 SEIPSSIEVLTSSGISSSVEPTSLVGPSSDESISSTESLSATSTFTSAVVSSSKAADFFT 652

Qy    916 -SDNVSKSGVSVTTETSVTTIQTTPN-PLSSSVTSLTQ----------LSSIPSVSESES 963
             |     :|| ||    :   |    ||     ||: ||: | | :|              |  |  |   || :
Db    653 RSTVSAKSDVSGNSSTQSTTFFATPSTPLAVSSTVVTSSTDSVSPNIPFSEISSSPESST 712

Qy    964 KVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNT-----REQ 1018
           :| ||         |            :|: :      | |:      |    :|||         :  |||::
Db    713 AITSTSTSFIAERTSSLYLSSSNMSSFTLSTFTVSQSIVSS---FSMEPTSSVASFASSS 769

Qy   1019 PTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTV-----ATSTLASASEEDNKSGSH 1073
           |   :|:  ||       |  :|         || |:       |:        |||| :  |   :
Db    770 PLLVSSRSN--CSDARSS--------NTISSGLFSTIENVRNATSTFTNLSTD------- 812

Qy   1074 ESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTT--EPTDTREQPTTL 1131
           |     || | |          : | |::      |: |          ||      ||      |    |
Db    813 EIVITSCKSSCTNEDSVLTKTQVSTVETTITSC------SGGICTTLMSPVTTINAKANT 866

Qy   1132 STTSKTNSELVATTQATNENGGKS----PSTDLTSSLTTGTSASTSANSELVTS------ 1181
                ||::|::   | ||   |     |      |  | :||   ||  :  |     | | :||
Db    867 LTTTETST--VETTITTCPGGVCSTLTVPVTTITSEATTTATISCEDNEEDITSTETELL 924

Qy   1182 ------GSVTGGAVASASN-----DQSHSTSVTNSNSIVSNTPQT-------TLSQQVTS 1223
                         | :||    :   :         :    :|   |      | | |    |             ||:  ||:
Db    925 TLETTITSCSGGICTTLMSPVTTINAKANTLTTTETSTVETTITTCSGGVCSTLTVPVTT 984

Qy   1224 -SSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLS 1257
               :| :| |     |    |         :        | |    :  |  | ::
Db    985 ITSEATTTATISCEDNEEDVASTKTELLTMETTIT 1019


RESULT 12
FIG2_YEAST
ID   FIG2_YEAST      STANDARD;      PRT;  1609 AA.
AC   P25653;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Factor induced gene 2.
GN   FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92397594; PubMed=1523889;
RA   Wilson C., Grisanti P., Frontali L.;
RT   "The complete sequence of a 6146 bp fragment of Saccharomyces
RT   cerevisiae chromosome III contains two new open reading frames.";
RL   Yeast 8:569-575(1992).
CC   -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC   -!- INDUCTION: BY MATING PHEROMONES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X59720; CAA42254.1; -.
DR   PIR; S19504; S19504.
DR   PIR; S25345; S25345.
DR   SGD; S0000685; FIG2.
SQ   SEQUENCE   1609 AA;  166049 MW;  7D66AD7F85A7B852 CRC64;

  Query Match             8.4%;  Score 547.5;  DB 1;  Length 1609;
  Best Local Similarity   22.5%;  Pred. No. 9.8e-17;
  Matches  332;  Conservative  232;  Mismatches  503;  Indels  411;  Gaps   66;

Qy     55 SLDGTSANPGDTFT-LNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCT 113
          ||  || ||   ::|  |          | :     :::| |:  :   :|       | | | |   |:| :
Db    165 SLTFTSVNPSQSWTSFNSE---KSSALSSTIDFTSSEISGSTSP-KSLESFDTTGTITSS 220

Qy    114 VNDA------------------------LKSSIKAF----GTVTLPIAFNVGGTGSST 143
           : :                         | |:|:|            : |:|    :     :     |
Db    221 YSPSPSSKNSNQTSLLSPLEPLSSSSGDLILSSTIQATTNDQTSKTIPTLVDATSSLPPT 280

Qy    144 DLEDSKCFTAGTNTVTFN--------DGDKDISIDVEFEKSTVDPSAYL----YASRVMP 191
                   |    |:|:::::: |                |: |:         :::|||  :       |:|   :
Db    281 LRSSSMAPTSGSDSISHNFTSPPSKTSGNYDV-----LTSNSIDPSLFTTTSEYSSTQLS 335

Qy    192 SLNKVT--------------------TLFVAPQCENGYT-SGTMGFSSSNGDVAIDCSN 229
           |||: :                          |   : |      | | |    :| |::|          :  :
Db    336 SLNRASKSETVNFTASIASTPFGTDSATSLIDPISSVGSTASSFVGISTANFSTQGNSNY 395
```

```
Qy    230 IHIGITKG---LNDW---NYPVS--------------------SESFSYTKTCTSNGIQI 263
          :     |     ||    : |:|                    | ::  | | | :|:
Db    396 VPESTASGSSQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVSTATKTVDGVIT 455

Qy    264 KY---------QNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTG 314
          :|         ::   |    | :   |:  :  ::  :|  | | |    : | :
Db    456 EYVTWCPLTQTKSQAIGVSSSISSVPQASSFSGSSILSSNSSTLAAS--NNVPESTASGS 513

Qy    315 YKNSDAGSNGIVIVATT-----RTVTDSTTAVTTLPFNPSVDKTKTI-----EILQPIPT 364
          :  |  |: : :  ||      | |  :   || |   |   |||:     | :   |
Db    514 SQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVST-ATKTVDGVITEYVTWCPL 572

Qy    365 TTITTSYVGVTTSYLTKT--------APIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
          |   :  :|:::| :: |          :| :   : |  :  | |:    |  :| :|:
Db    573 TQTKSQAIGISSSTISATQTSKPSSILTLGISTLQLSDATFKGTETIN---THLMTESTS 629

Qy    417 RTNPT------DSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPP 470
          | ||       ||      :| | : ::|:   :| |  :| |:|
Db    630 ITEPTYFSGTSDSFYLCTSEVNLAS-SLSSYPNFSSSEGSTATIT--------------- 673

Qy    471 NHTVT---TTEYWSQSFATTTTVTA-----------PPGGTDSVIIREPPNPTVTTTEY 515
          | |||   |::| | | :  |  :                |::: :    : | :  :|
Db    674 NSTVTFGSTSKYPSTSVSNPTEASQHVSSSVNSLTDFTSNSTETIAVISNIHKTSSNKDY 733

Qy    516 WSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT------TEY--W--SQSYATTTTVTAP 565
            : ||       |  ::::      ||||      |||  |  : | | ||:::
Db    734 ------SLTTTQLKTSGKQTLVL-----STVTTTVNGAATEYTTWCPASSIAYTTSISY- 781

Qy    566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTV---TTTE 622
                    | |||  | |  | | :|:    : |: :    : ||   | :|
Db    782 -------------KTLVLTTEVCSHSECTPTVITSVTATSSTIPLLSTSSSTVLSSTVSE 828

Qy    623 YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI 682
              |:  |:    | |          | |:|:       : |: ||   |: |
Db    829 GAKNPAASEVTINTQVSATS--------EATSTSTQ------VSATSATATASESSTTSQ 874

Qy    683 REPPNHTVTTTEYWSQSYATTTTVTAPPGETD----TVLIREPPNHTVTTTEYWSQSYAT 738
            : |::|    :|:: || ::  |   |:    ||: |:     : :||  | |
Db    875 VSTASETIST--LGTQNFTTTGSLLFPALSTEMINTTVVSRKT---LIISTEVCSHSKCV 929

Qy    739 TTTVT-------APPGG-------TDTVIIREPPNPTVT-TTEYWSQSFATTTTVTAPPG 783
            | :|         |  |        |: | :    : ||| :||  | |  | | :|:
Db    930 PTVITEVVTSKGTPSNGHSSQTLQTEAVEVTLSSHQTVTMSTEVCSNSICTPTVITSVQ- 988

Qy    784 GTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSD 843
            |   | : |:|  | :|   :|:             :|::|||  |: : | : :
Db    989 MRSTPFFPYLTSSTSSSSLASTKKSSL-----------EASSEMSTFSVSTQSLPLAFTCS 1037

      844 GMLLSSTTLVTESETT--TELICSDGKECSRLSSSSGIVTN--PDSNES--------SIV 891
              |||  |::   |  |  | |         |||: | ||  | |  |         |:
     1038 EK--RSTTSVSQWSNTVLTNTIMS--------SSSNVISTNEKPSSTTSPYNFSSGYSLP 1087

Qy    892 TSTVPTASTMSDSLSSTDGI-----------------------------SATSSDNVS 920
          :|: |:  ::| : :: :||                             |:  | ::|
Db   1088 SSSTPSQYSLSTATTTINGIKTVYTTWCPLAEKSTVAASSQSSRSVDRFVSSSKPSSSLS 1147

Qy    921 KSGVSVTTETSVTTIQ------TTPNPLSSSVT--SLTQLSSIPSVSESESKVTFTSNGD 972
          :: :  |  |: |||       ||  ||:|  |  : || |        | :||  ||
Db   1148 QTSIQYTLSTATTTISGLKTVYTTWCPLTSKSTLGATTQTS-------STAKVRITS-AS 1199

Qy    973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE-PTNTREQPTTLS-----TTS 1026
          : : |  | |||||  |    ||:  |   | | |: | : || :    :||:      :||
Db   1200 SATSTSISLSTSTESE----SSSGYLSKGVCSGTECTQDVPTQSSSPASTLAYSPSVSTS 1255

Qy   1027 NSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGE 1086
          :| :   ||:          ||:   || : :|:  |||   : |      ||||
Db   1256 SSSSFSTTTASTL------TSTHTSVPLLPSSSSISASSPSSTS----LLSTSLPSPAFT 1305

Qy   1087 NSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSEL----- 1141
          :| | |:| : :||    :  |   :|| : ::  |  :    :  |::| ::| |
Db   1306 SSTLPTATAVSSSTFIASSLP---LSSKSSLSLSPVSSSILMSQFSSSSSSSSSLASLPS 1362

Qy   1142 -----------------VATTQATNENGGKSPST--------DLTSSLT------TGTS 1169
```

```
                              :||   |:    : ||        |:||: |      | |
Db   1363 LSISPTVDTVSVLQPTTSIATLTCTDSQCQQEVSTICNGSNCDDVTSTATTPPSTVTDTM 1422

Qy   1170 ASTSANSELVTSGSVTG-----------GAVASASND--QSHSTSVTNS-----NSIVS 1210
                | :  :  || |  |              ::: | :  |: :||  ||      |:::
Db   1423 TCTGSECQKTTSSSCDGYSCKVSETYKSSATISACSGEGCQASATSELNSQYVTMTSVIT 1482

Qy   1211 NTPQTTLSQQV--TSSSPSTNTFIASTY---DGSGSII 1243
               :  || | :|  | |: |  |    ||   | :| :|
Db   1483 PSAITTTSVEVHSTESTISITTVKPVTYTSSDTNGELI 1520


RESULT 13
YM96_YEAST
ID   YM96_YEAST     STANDARD;      PRT;  1140 AA.
AC   Q04893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN   YMR317W OR YM9924.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z54141; CAA90835.1; -.
DR   SGD; S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE   1140 AA;  113070 MW;  0153EBCA24FE5427 CRC64;


  Query Match             8.0%;  Score 521.5;  DB 1;  Length 1140;
  Best Local Similarity   24.1%;  Pred. No. 8.8e-16;
  Matches  243;  Conservative  193;  Mismatches  379;  Indels  195;  Gaps   39;

Qy    320 AGSNGIVIVATTR---TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 376
          :||       |:    | | |||: ||                    |||| |||    ||
Db      4 SGSKSTTATTTSHSSTTTTSSTTSTTT------------------PTTTSTTS----TT 40

Qy    377 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNP 436
          |   |:|      :||      ::|  |:| : :  :    :|:| :::     | :
Db     41 STKVTTSP-----EIIVSSSSTLVSSVVPEFTSSSSLS------SDTIASILSSESLVS- 88

Qy    437 TVSTTEYWSQSFATTT-----TVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVT 491
            |:  | |   ::|:     : |: |  : :, :        : :|||  : | :::   |
Db     89 IFSSLSYTSSDISSTSVNDVESSTSGPSNSYSALSSTNAQLSSSTTE--TDSISSSAIQT 146

Qy    492 APP-------GGTDS-----VIIREPPNPTVTTTEYWSQSFATTTTVTAPP--GGTDSVI 537
          : |       ||: |       : |    :  ||   | :| | | |:: |    : | :
Db    147 SSPQTSSSNGGGSSSEPLGKSSVLETTASSSDTTAVTSSTFTTLTDVSSSPKISSSGSAV 206

Qy    538 IREPPNPTVTTTEYWSQSY--ATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATT 595
                :| ||   |:    ::|: |::    | |     | : |:: |  :|   :|
Db    207 ------TSVGTTSDASKEVFSSSTSDVSSLLSSTSS-----PASSTISETLPFS---STI 252

Qy    596 TTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVT 655
          ::|: |       :  | |: | ::    : |  :::   :   |  | :|:  | |:
Db    253 LSITSSP------VSSEAPSATSSSVSSEASSSTSSSVSSEAPLATSSVVSSEAPS---- 302
```

```
Qy    656 TTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDT 715
Db    303 -------------------STSSVVSSEAPSSTSSSVSSEISSTTSSSVSSEAPLATSS 342

Qy    716 VLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATT 775
Db    343 VVSSEAPSSTSSSVSSEISSTTSSSVSSEAPLATSSVVSSEAPSSTSSSVSSEAPSSTSS 402

Qy    776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMN 835
Db    403 SVSSEAPSSTSSSVSSE-ISSTKSSVMSSE------------VSSATSSLVSSEAPS-- 446

Qy    836 TPTSISSDGMLLSSTTLVTESETTTEL--ICSDGKECSRLSSSSGIVTNPDSN--ESSIV 891
Db    447 AISSLASSRLFSSKNTSVTSTLVATEASSVTSSLRPSSETLASNSIIESSLSTGYNSTVS 506

Qy    892 TSTVPTASTMSDSLSSTDGISATS-----SDNVSKSGVSVTTETSVTT-----IQTTPNP 941
Db    507 TTTSAASSTLGSKVSSSNSRMATSKTSSTSSDLSKSSVIFGNSSTVTTSPSASISLTASP 566

Qy    942 LSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE-IVTTSSTKVLPP 1000
Db    567 LPSVWSDITS-------SEASS---ISSNLASSSAPSDNNSTIASASLIVTKTKNSVVSS 616

Qy   1001 VVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTL 1060
Db    617 IVSSIT--SSETTN----ESNLATSSTSLLSNKATARSLSTS--NATSASNVPTGTFSSM 668

Qy   1061 ASASEEDNKSGSHESASTSLKPSMGENS--GLTTSTEIEATTTS---PTEAPSPAVSSGT 1115
Db    669 SSHTSVITPGFSTSSASLAINSTVVSSSLAGYSFSTPESSPTTSTLVTSEAPS-TVSSMT 727

Qy   1116 DVTTEP-----TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSA 1170
Db    728 --TSAPFINNSTSARPSPSTASFITESTSSISSVPLASGDVTSSLAAHNLTTFSAPSTSS 785

Qy   1171 ------STSANSELVT-----SGSVTGGA------------VASASNDQSHSTSVTNSNS 1207
Db    786 AQLVSKSTTSSSILVTPRIDRSGNSSTASRIATSLPNKTTFVSSLSSTSAHARNIFNSTV 845

Qy   1208 IVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLS 1257
Db    846 LATAKQIETLTSTVNCSNPTPNYNITKTV-----IVSRETTAIGTVTSCS 890


RESULT  14
YK82_YEAST
ID   YK82_YEAST     STANDARD;      PRT;  1169 AA.
AC   P36170;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Hypothetical 122.2 kDa protein in SIR1 3'region precursor.
GN   YKR102W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gaillon L., Dujon B.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z28327; CAA82182.1; -.
DR   PIR; S38181; S38181.
DR   SGD; S0001810; YKR102W.
KW   Hypothetical protein; Glycoprotein; Repeat; Cell wall; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27   1169       HYPOTHETICAL PROTEIN YKR102W.
FT   DOMAIN      274    609       SER/THR-RICH.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    389    389       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    452    452       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    879    879       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1092   1092       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1099   1099       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1169 AA;  122164 MW;  940C6DFB3569C669 CRC64;


  Query Match             7.7%;  Score 501;  DB 1;  Length 1169;
  Best Local Similarity   23.1%;  Pred. No. 6.9e-15;
  Matches  313;  Conservative  208;  Mismatches  528;  Indels  308;  Gaps   53;

Qy      6 TLLFLYLSIASAKTIT------------GVFDSFNSLTWSNAANYAFKGPGYPTWNAVLG 53
Db     11 TGLFL-LSVANVALGTTEACLPAGEKKNGMTINFYQYSLKDSSTYS--NPSYMAYGYADA 67

Qy     54 WSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGE-------EFTT 106
Db     68 EKLGSVSGQTKLSIDYSIPCNGASDTCACSDD---DATEYSASQVVPVKRGVKLCSDNTT 124

Qy    107 FSTLT-------CTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVT 159
Db    125 LSSKTEKRENDDCDQGAAYWSS-DLFGFYTTPTNVTVEMTG--------YFLPPKTGTYT 175

Qy    160 FNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSS 219
Db    176 FG-------------FATVDDSAIL---------------------------------SV 189

Qy    220 NGDVAIDCSNIH---IGIT----KGLNDWN--YPVSSESFSYTKTCTSNGIQIKYQN-VP 269
Db    190 GGNVAFECCKQEQPPITSTDFTINGIKPWNADAPTDIKGSTYMYAGYYYPIKIVYSNAVS 249

Qy    270 AGYRPFIDAYISATDVNQ----YTLAYTND--------------YTCAGSRLQSKPFTL 310
Db    250 WGTLPVSVVLPDGTEVNDDFEGYVFSFDDNATQAHCSVPNPAEHARTCVSSATSS----- 304

Qy    311 RWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTS 370
Db    305 -WSSSE-----------VCTECTETESTSYVTPYVTSSSWSSSEVCTECTETESTSTSTP 352

Qy    371 YVGVTTSYLTKT-APIGETATVIVDVPYHTTTTVTSEWTGTITTTT---TRTNPTDSIDT 426
Db    353 YVTSSSSSSSEVCTECTETESTSYVTPYVSSSTAAANYTSSFSSSSEVCTECTETESTST 412

Qy    427 VVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486
Db    413 STPYV--------TSSSWSSSEVCTECTE-----TESTSYVTPYVSSSTAAANYTSSFSS 459

Qy    487 TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTV 546
Db    460 SSEVCTECTETES---------TSTSTPYVTSSSSSSSEVCTECTETESTSYVTPYVSSS 510

Qy    547 TTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD 606
Db    511 TAAANYTSSFSSSSEVCTECTETES---------TSTSTPYVTSSSWSSSEVCTECTETE 561

Qy    607 TVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYAT 666
```

```
Db    562 STSYVTPYVSSSTAAANYTSSFSSSSEVCTECTETES---------TSTSTPYATSSTGT 612

Qy    667 TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTV 726

Db    613 ATSFTASTSNTMTSLVQ------TDTTVSFSLSSTVSEHTNAP-----TSSVESNASTFI 661

Qy    727 TTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVT-----TTEYWSQSFATTTTVTAP 781

Db    662 SSNKGSVKSYVTSSIHSITP--------MYPSNQTVTSSSVVSTPITSESSESSASVTIL 713

Qy    782 PGGTDTVIIYESMSSSKISTSSNDITSIIPSF---SRPHYVNSTTSDLSTFESSSMNTPT 838

Db    714 P-STITSEFKPSTMKTKVVSISSSPTNLITSYDTTSKDSTVGSSTSSVSLISSISLPSSY 772

Qy    839 SISSD----GMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTST 894

Db    773 SASSEQIFHSSIVSSNGQALTSFSSTKVSSSESSESHRTSPTT-------SSESGIKSSG 825

Qy    895 VPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSS 954

Db    826 VEIESTSTSSFSFHE--TSTASTSVQISSQFVTPSSPISTVA----PRSTGLNSQTE--- 876

Qy    955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTD------- 1007

Db    877 ----STNSSKETMSSE-NSASVMPSSSATSPKTGKVTSDET-------SSGFSRDRTTVY 924

Qy   1008 -LTSEPTNTREQPTTLSTT-------SNSITEDITTSQPTGDNGDNTSSTNPVPTVATST 1059

Db    925 RMTSETPSTNEQTTLITVSSCESNSCSNTVSSAVVSTATTTINGITTEYTTWCPLSATE- 983

Qy   1060 LASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP---SPAVSSGTD 1116

Db    984 LTTVSKLESE---------------EKTTLITVTSCESGVCSETASPAIVSTATATVND 1027

Qy   1117 VTTEPTDTREQPTT-LSTTSKTNSEL--------VATTQATNENGGKSPSTDLTSSLT-T 1166

Db   1028 VVTVYSTWSPQATNKLAVSSDIENSASKASFVSEAAETKSISRNNNFVPTSGTTSIETHT 1087

Qy   1167 GTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSI-VSNTPQTTLSQQVTSSS 1225

Db   1088 TTTSNASENSDNV-----------SASEAVS-SKSVTNPVLISVSQQPRGTPASSMIGS- 1134

Qy   1226 PSTNTFIASTYDGSGSIIQHSTWLYGLITL--LSLFI 1260

Db   1135 -STASLEMSSYLG---IANH------LLTNSGISIFI 1161


RESULT 15
VG50_HSVI1
     VG50_HSVI1     STANDARD;      PRT;   670 AA.
AC   Q00130;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical gene 50 protein.
GN   50.
OS   Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   unclassified Herpesviridae.
OX   NCBI_TaxID=10401;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AUBURN 1;
RX   MEDLINE=92087490; PubMed=1727613;
RA   Davison A.J.;
RT   "Channel catfish virus: a new type of herpesvirus.";
RL   Virology 186:9-14(1992).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M75136; AAA88153.1; -.
DR   PIR; F36791; F36791.
KW   Hypothetical protein; Repeat.
FT   REPEAT      143    158       1.
FT   REPEAT      171    186       2.
FT   REPEAT      200    214       3.
FT   REPEAT      215    233       4.
FT   REPEAT      234    252       5.
FT   REPEAT      253    268       6.
FT   REPEAT      279    293       7.
FT   REPEAT      294    309       8.
FT   REPEAT      320    334       9.
FT   REPEAT      335    349       10.
FT   REPEAT      362    376       11.
FT   REPEAT      377    391       12.
FT   REPEAT      392    406       13.
FT   REPEAT      407    421       14.
FT   REPEAT      422    436       15.
FT   REPEAT      437    452       16.
FT   REPEAT      464    477       17.
FT   REPEAT      478    493       18.
FT   REPEAT      504    517       19.
FT   REPEAT      518    531       20.
FT   REPEAT      532    545       21.
FT   REPEAT      546    559       22.
FT   REPEAT      560    573       23.
FT   REPEAT      574    587       24.
FT   REPEAT      588    601       25.
FT   REPEAT      602    615       26.
FT   REPEAT      616    629       27.
SQ   SEQUENCE   670 AA;  64174 MW;  2B64A781C519E8B4 CRC64;


  Query Match             7.2%;  Score 464.5;  DB 1;  Length 670;
  Best Local Similarity   24.3%;  Pred. No. 1.4e-13;
  Matches  215;  Conservative   71;  Mismatches  294;  Indels  303;  Gaps   34;

Qy    297 TCAGSRLQSKPFTLRWT--GYKNS--DAGSNGIVIVATTRTVTDST-----------TAV 341

Db      9 TCLGMSLLASNLALVGVVRSYTNETPTPGPETELITTTVARMTDPTAKPSDFPGDAVTGT 68

Qy    342 TTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAP---IGETATVI--VDVP 396

Db     69 QPVPREPS----------SLPRTTPSLAHTTISKMISLGTRPRPTIPGVPTTIPNTDAP 117

Qy    397 -----YHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATT 451

Db    118 VDPGSVHTTARVVTDIT-TKQTPTTPATPAGANDTANITTATPAGANDT--------ANI 168

Qy    452 TTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVT 511

Db    169 TTAT-PAGANDTANI-------TTATPAGANDTAVTTTSATPAGANDT------------ 208

Qy    512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDS 571

Db    209 ---------AVTTTPATPAGANDTA------NDTV-----------VTTTPAMPAGANDT 242

Qy    572 VIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT 631

Db    243 ----------------ANGTAVTTTPAMPAGANDTANIT-----TATPT-----GANDT 275

Qy    632 TTVTGPPSGTDTVIIREPPNPT-------VTTTEYWSQSYAT------TTTITAPPGETD 678

Db    276 ANVTMPAGATDTVVTTTPAMPTGANDTANITT--------ATPAGANDTANVTMPAGATD 327

Qy    679 TVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYAT 738

Db    328 TV---------VTTTPAMPAGANDTANVTKPAGSTDTVV---------------------- 357
```

```
Qy    739 TTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSK 798
          |||   | | ||||:   |  ||  |          |||   | | ||||:
Db    358 TTTPAMPTGATDTVVTTTPAMPTGAT------DTVVTTTPAMPTGATDTVV--------- 402

Qy    799 ISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESET 858
                                                                       |
Db    403 ----------------------------------------------------------T 403

Qy    859 TTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSST-DGISATSSD 917
          ||                    |     ::||:|    :  :|:: :|     | ::|
Db    404 TTP------------------AKPAGANGTVVTTTPAMPAGANDTVVTTAPATPAGAND 444

Qy    918 NVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGT 977
            : :  :  |:| |||    |   : :||: |               |   |  : |
Db    445 TANVTKPTGATDTVVTTATVKPTGATGTVTTTTAKP-------------TGANDTANVT 490

Qy    978 HDSQSTSTEIEIVTTSSTKVLPPVVSSN--TDLTSEPTNTREQPTTLSTTSNSITEDITT 1035
            : :| |    |||:: |   |  ::   |  |::||      ||            ||
Db    491 KPTGATGT----VTTTTAK---PTGATGTVTVATAKPTGATGTVTT-----------TT 531

Qy   1036 SQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTST- 1094
          ::||| ||  |::|   || || |: : : :   :|:: : :|:     | |  :||:|
Db    532 AKPTGANGTVTTTT-AKPTGATGTVTTTTAK--PTGANGTVTTTTAKPAGANGTVTTTTA 588

Qy   1095 -EIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSK 1136
             |  |  |    || ::|| ||         |  :||:|
Db    589 KPAGANGTVTTTTAKPAGANGTVTTTTAKPAGANGTVTTTTAK 631
```

Search completed: October 3, 2002, 15:37:58
Job time: 324 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:32:19 ; Search time 46.21 Seconds
(without alignments)
4717.023 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

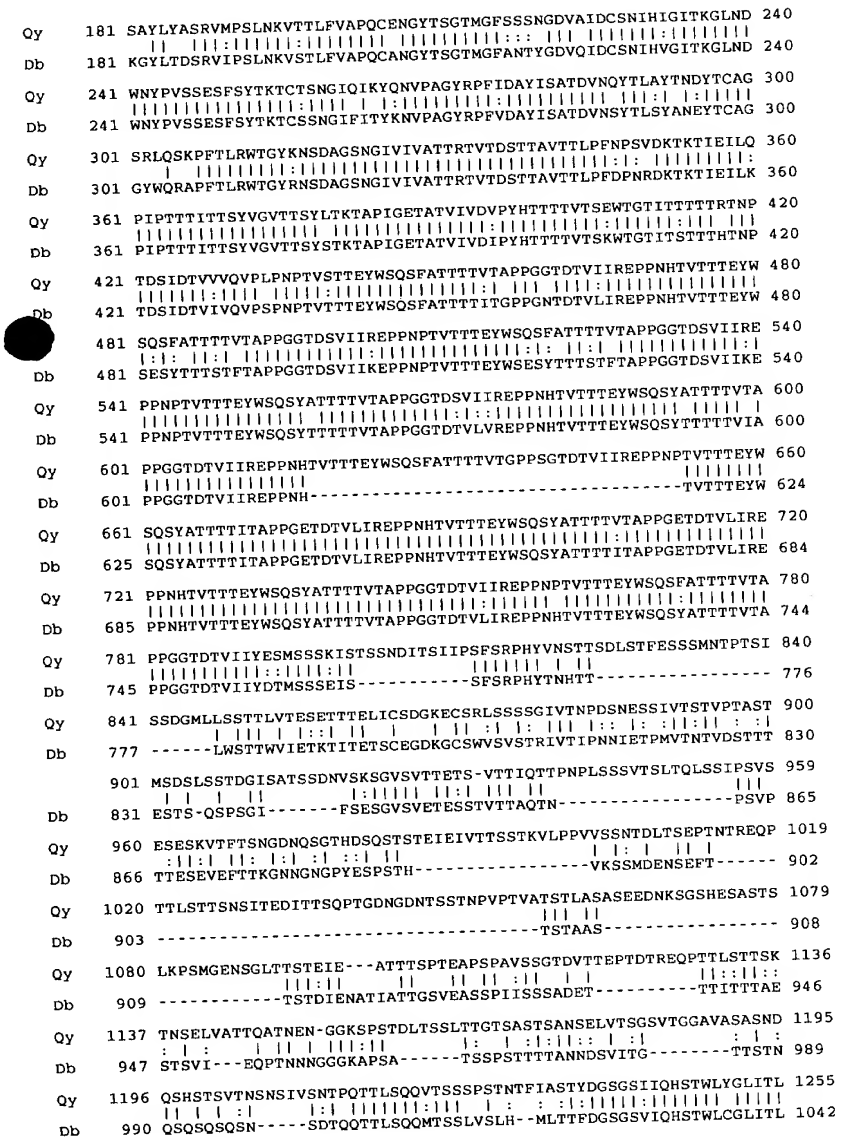
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4078.5	62.8	1047	3	O93865	O93865 candida alb
2	3716.5	57.2	1523	3	Q9HFX4	Q9hfx4 candida alb
3	2791.5	43.0	1443	3	Q9Y8F2	Q9y8f2 candida alb
4	2673.5	41.2	1270	3	Q9Y743	Q9y743 candida alb
5	2576.5	39.7	2297	3	Q9HGK6	Q9hgk6 candida alb
6	1986.5	30.6	468	3	Q9URQ0	Q9urq0 candida alb
7	1823	28.1	469	3	Q9URP8	Q9urp8 candida alb
8	1780.5	27.4	468	3	Q9C471	Q9c471 candida alb
9	1471	22.6	336	3	Q9HF70	Q9hf70 candida dub
10	1191	18.3	338	3	Q9HF72	Q9hf72 candida dub
11	1176	18.1	331	3	Q9HF69	Q9hf69 candida dub
12	999.5	15.4	433	3	Q9Y8F1	Q9y8f1 candida alb
13	981	15.1	353	3	Q9HF71	Q9hf71 candida tro
14	831.5	12.8	1195	3	Q96WU8	Q96wu8 schizosacch
15	824	12.7	226	3	Q9HG16	Q9hg16 candida alb
16	698.5	10.8	1079	5	Q9N4S7	Q9n4s7 caenorhabdi
17	685	10.5	2586	5	Q9VTK8	Q9vtk8 drosophila
18	674.5	10.4	1220	3	Q9C0Y2	Q9c0y2 schizosacch
19	649.5	10.0	1275	5	O76602	O76602 caenorhabdi
20	646.5	10.0	2232	5	P91365	P91365 caenorhabdi
21	637.5	9.8	2271	16	Q99QY4	Q99qy4 staphylococ
22	632.5	9.7	3570	4	Q99552	Q99552 homo sapien
23	620	9.5	3072	2	Q939N5	Q939n5 streptococc
24	601	9.3	1795	5	O76894	O76894 drosophila
25	600.5	9.2	2178	2	Q9KWR3	Q9kwr3 streptococc
26	594.5	9.2	3971	3	Q96WV6	Q96wv6 schizosacch
27	582.5	9.0	1752	2	Q9AE52	Q9ae52 ruminococcu
28	578	8.9	1283	3	Q9USQ3	Q9usq3 schizosacch
29	577.5	8.9	957	4	O14651	O14651 homo sapien
30	577.5	8.9	1217	4	Q9UKW9	Q9ukw9 homo sapien
31	574	8.8	1283	3	Q9URU4	Q9uru4 schizosacch
32	570.5	8.8	858	3	Q9C470	Q9c470 candida alb
33	553.5	8.5	1459	5	O17084	O17084 caenorhabdi
34	553.5	8.5	3178	5	Q969D4	Q969d4 caenorhabdi
35	545.5	8.4	957	4	Q9UKN0	Q9ukn0 homo sapien
36	538.5	8.3	406	3	Q9URP9	Q9urp9 candida alb
37	538.5	8.3	1203	5	Q9N5K0	Q9n5k0 caenorhabdi
38	534	8.2	849	3	P87107	P87107 saccharomyc
39	533	8.2	407	3	O74661	O74661 candida alb
40	531.5	8.2	406	3	O74658	O74658 candida alb
41	529	8.1	407	3	O74659	O74659 candida alb
42	526	8.1	1229	5	Q94185	Q94185 caenorhabdi
43	515	7.9	4776	16	Q97P71	Q97p71 streptococc
44	511.5	7.9	3507	5	Q23587	Q23587 caenorhabdi
45	503.5	7.8	1844	5	Q22579	Q22579 caenorhabdi

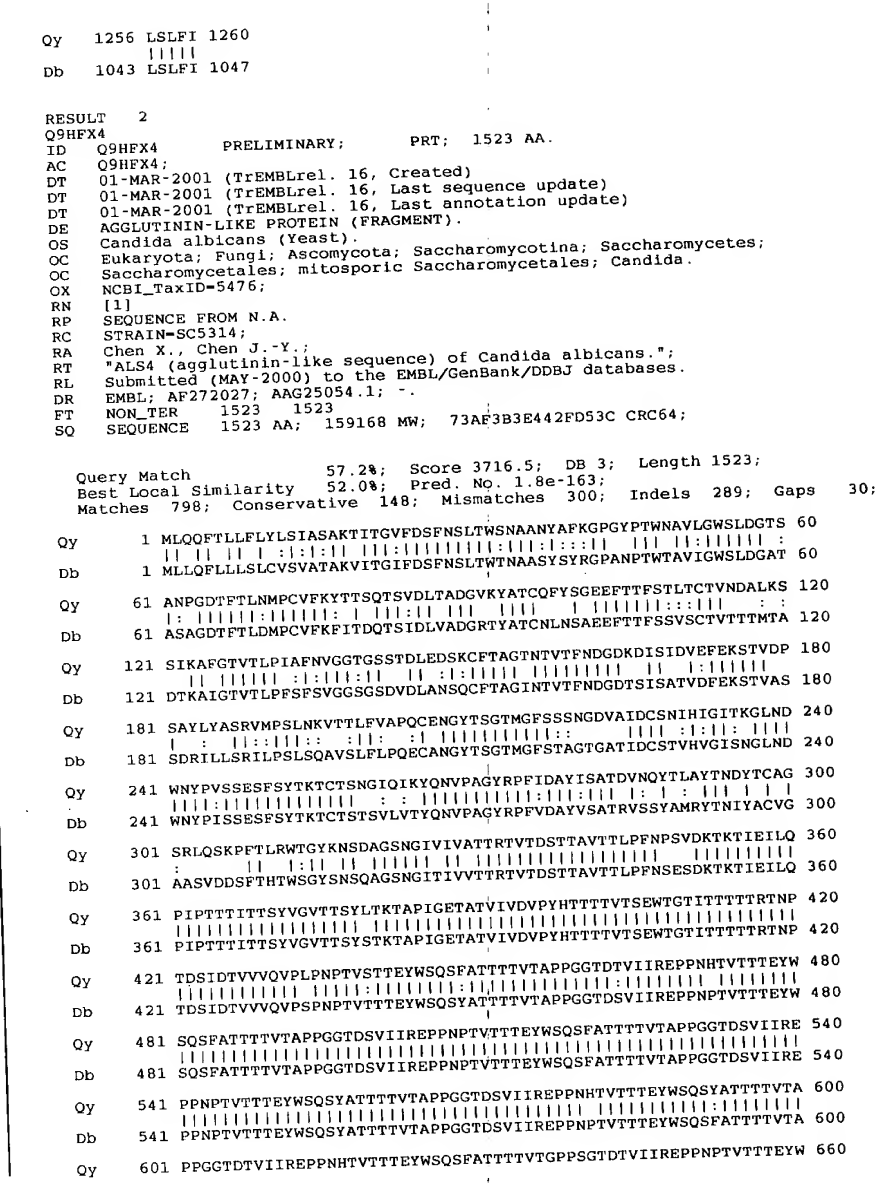
ALIGNMENTS

```
RESULT   1
O93865
ID   O93865       PRELIMINARY;       PRT;  1047 AA.
AC   O93865;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE CELL SURFACE PROTEIN.
GN   ALS8.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC10261;
RA   Leng P., Lee P.R., Wishart J.A., Wu H., Brown A.J.P.;
RT   "Sequence of the hypha-specific, agglutinin-like cell surface protein,
RT   ALS8 from Candida albicans.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF051313; AAD02580.1; -.
SQ   SEQUENCE   1047 AA;  111944 MW;  C2327659AA911F2E CRC64;
```

Query Match 62.8%; Score 4078.5; DB 3; Length 1047;
Best Local Similarity 65.0%; Pred. No. 2.7e-180;
Matches 822; Conservative 100; Mismatches 120; Indels 223; Gaps 18;

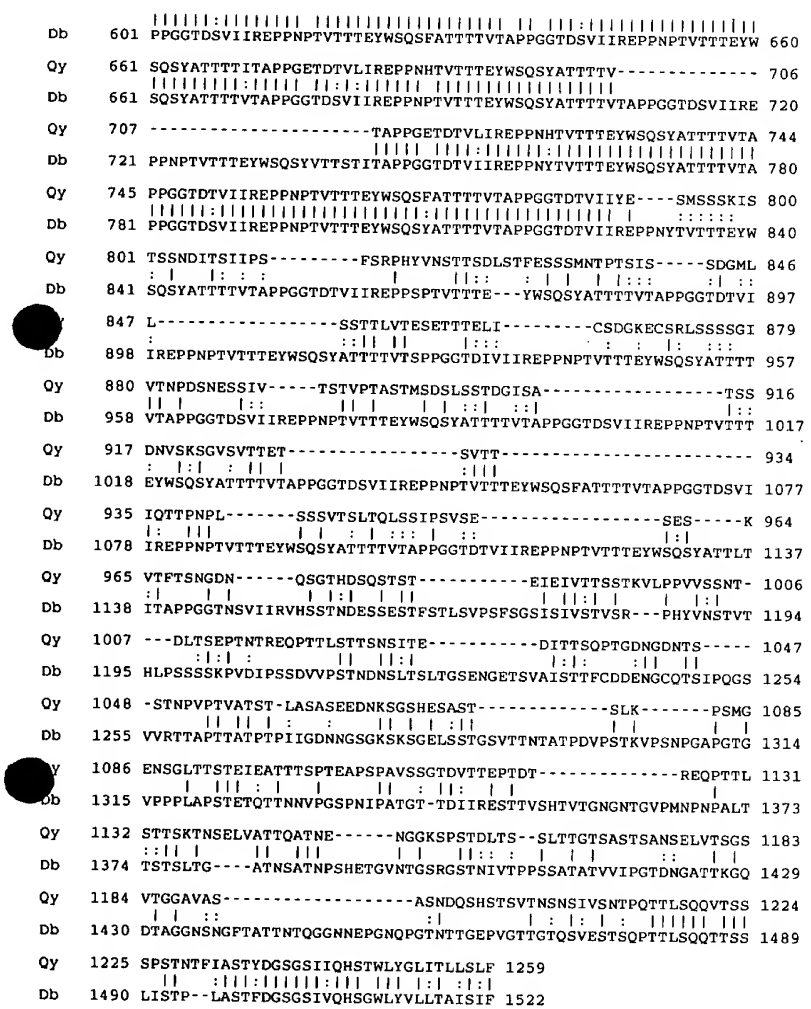
```
Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           ||||:||| :|||:|:|||||||:||||||||||| | :|||| ||||||||||||||
Db       1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYHYKGPGTPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           |:|||||||||||||:|||||||||| |||||||||| :|||| ||||||||:: |
Db      61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           |||| |||||:|||||||||| |||||||||||||||||||| | |||:|:||:| |||
Db     121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180
```





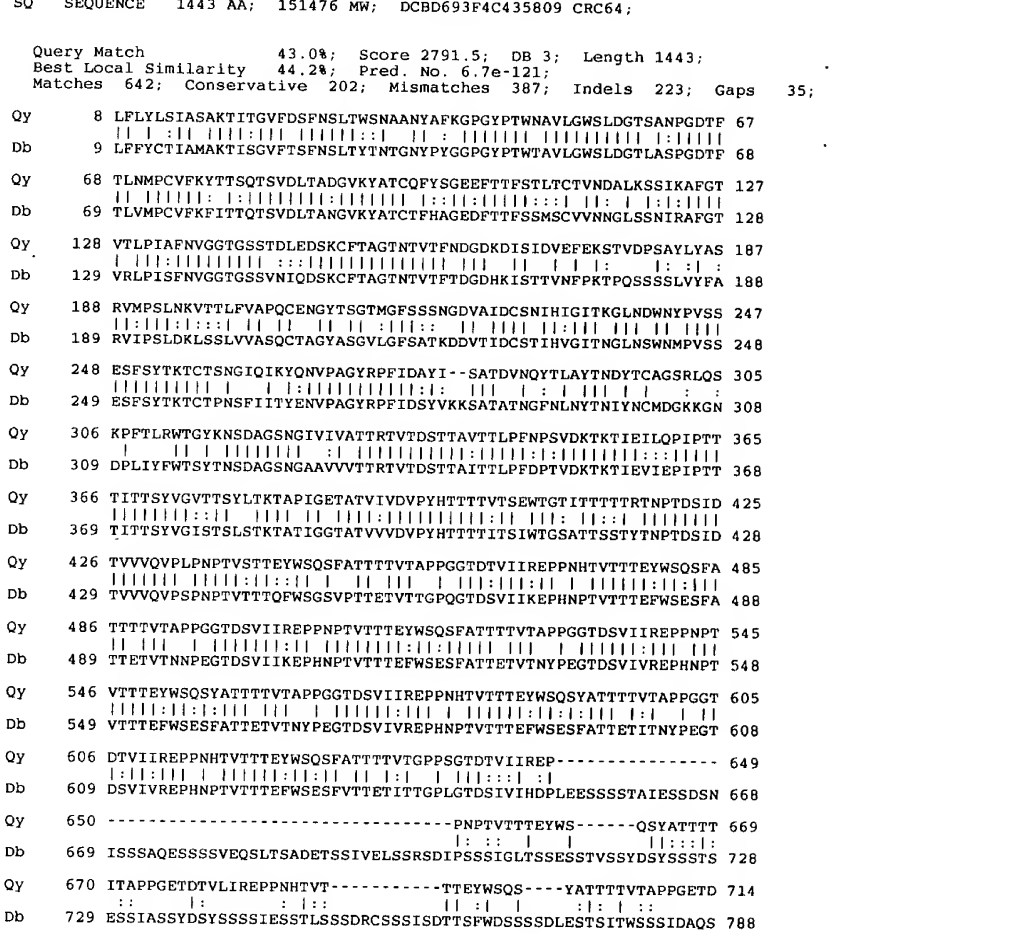
RESULT 2
Q9HFX4
ID Q9HFX4 PRELIMINARY; PRT; 1523 AA.
AC Q9HFX4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE AGGLUTININ-LIKE PROTEIN (FRAGMENT).
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RA Chen X., Chen J.-Y.;
RT "ALS4 (agglutinin-like sequence) of Candida albicans.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF272027; AAG25054.1; -.
FT NON_TER 1523 1523
SQ SEQUENCE 1523 AA; 159168 MW; 73AF3B3E442FD53C CRC64;

Query Match 57.2%; Score 3716.5; DB 3; Length 1523;
Best Local Similarity 52.0%; Pred. No. 1.8e-163;
Matches 798; Conservative 148; Mismatches 300; Indels 289; Gaps 30;



RESULT 3
Q9Y8F2
ID Q9Y8F2 PRELIMINARY; PRT; 1443 AA.
AC Q9Y8F2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AGGLUTININ-LIKE PROTEIN 6.
GN ALS6.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=20321177; PubMed=10861907;
RA Hoyer L.L., Hecht J.E.;
RT "The ALS6 and ALS7 genes of Candida albicans.";
RL Yeast 16:847-855(2000).
DR EMBL; AF075293; AAD42033.1; -.
DR InterPro; IPR003015; HLH_Myc.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 1443 AA; 151476 MW; DCBD693F4C435809 CRC64;

Query Match 43.0%; Score 2791.5; DB 3; Length 1443;
Best Local Similarity 44.2%; Pred. No. 6.7e-121;
Matches 642; Conservative 202; Mismatches 387; Indels 223; Gaps 35;



```
Qy    715 TVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREP------------PNPTV 762
Db    789 SHLVQSVSNSISTSQELSSSSSEESSTF-----ATDALVSSDASSILSSDTSSYYPSSTI 843

Qy    763 TTTEYW-----------SQSFATTT--------TVTAPPGGT-DTVIIYESMSSSKISTS 802
Db    844 SSSDDFPHTIAGESDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSD 903

Qy    803 SNDI-------TSIIPSFS----------RPHYVNSTTSDLSTFESSSMNTPTSISSDGM 845
Db    904 QSDILTSSSFSTLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESAN 963

Qy    846 LLSSTTLVTESETTTELICSDGKE---CSRLSSSSGIVTN-PDSNESSIVTSTVPTASTM 901
Db    964 --DDTYTLSESTDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSF 1021

Qy    902 SDSLSSTDG-ISATSSDNV--SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSV 958
Db   1022 VASSTPTSAEQSITDNPNIDSSQTSASSSTKLSVSVSDTVVNSISLSETS--TLSSDDST 1079

Qy    959 SESESKVTFTSNGD----NQSGTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSE 1011
Db   1080 S-SDTSISSTTNSDTGNVNAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITTE 1138

Qy   1012 PTNTREQPTTLSTTSNSITED----ITTSQP-------TGDNG------DNTSSTNPVPT 1054
Db   1139 LITTELITTELTTTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDGESTVEIPPV 1198

Qy   1055 VATSTLASASEEDNKSGSHESASTSLKPSMGENS-------GLTTSTEIEATTTSPTEA- 1106
Db   1199 TDTTTVSSVSVHSTE------ASTA---TLGENSFSKVASAPVNTETSLRSTSSSSNHAT 1249

Qy   1107 ---------------PSPAVSSGT--DVTTEPTDTREQPTTLSTTS-KTNSELVATTQAT 1148
Db   1250 ESSGTVKSEASAEAIPSPPTSTDNRLSYSTEEAKGSTYPNSGSTNNLMTESQVAAPTDST 1309

Qy   1149 N---ENGGKSPSTDLTSSLTTGTSASTSANSELVTS----GSVTGGAVASASNDQS---- 1197
Db   1310 SVLTANPVVTSTFDDKSSAAVNQPSKTKSIEESIGSLDSVNETNNGFIATLSQSEAPNSL 1369

Qy   1198 -HSTSVT-----------NSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQH 1245
Db   1370 IHSESISTTMAKTTDASINGDSAASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKH 1427

Qy   1246 STWLYGLITLLSLF 1259
Db   1428 PSWLLKFISVALFF 1441
```

```
RESULT 4
Q9Y743
ID   Q9Y743      PRELIMINARY;      PRT;  1270 AA.
AC   Q9Y743;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN.
GN   ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=21064501; PubMed=11124701;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS5 gene of Candida albicans and analysis of the Als5p N-
RT   terminal domain.";
RL   Yeast 18:49-60(2001).
DR   EMBL; AF068866; AAD32849.1; -.
SQ   SEQUENCE   1270 AA;  133337 MW;  D8E5FA5853F6D5C5 CRC64;
```

```
  Query Match             41.2%;  Score 2673.5;  DB 3;  Length 1270;
  Best Local Similarity   45.9%;  Pred. No. 1.6e-115;
  Matches  629;  Conservative  187;  Mismatches  340;  Indels  215;  Gaps   39;

Qy      1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db      1 MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTS 60

Qy     61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db     61 ANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy    121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db    121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGINTVTFNDGSKKLSIAVNFEKSTVDR 180

Qy    181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db    181 SGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVND 240

Qy    241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db    241 WNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVD 300

Qy    301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db    301 DYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy    361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db    361 PIPTTTITTSYVGVTTSYSTKTAPIGETATLIVDVPYHTTTTVTSKWTGTITMTTTRTNP 420

Qy    421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db    421 TDSIDTVVVQVPLPNPTTTTTQFWSESFTSTTTITNKPEGTDSVIVKEPHNPTVTTTEFW 480

Qy    481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTT-------------EYWSQSFATTTTVT 527
Db    481 SESYATTETITTGPLGTDSIVIHDPLEESSSTTAIESSDSNISSSAQESSSSVEQSSSIV 540

Qy    528 APPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEY 587
Db    541 GLSSSSDIPLSSDMPS-------------SSSTGLTSSESSTVSSYDSDSSSSSELSTFS 587

Qy    588 WSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW--SQSFATTTTVTGPPSGTDTVI 645
Db    588 SSESYSSSIS---------------------DTTNFWDSSSSDLESTSITWSSS------ 620

Qy    646 IREPPNPTVTTTEYWSQSYATTTTITAPPGE------TDTVLIREPPNHTVTTTEYWSQS 699
Db    621 IDAQSSQSVQSV---SNSISTSQETTSSSGEESNTSVTD-ILVSSDASSILNSD---ISS 673

Qy    700 YATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPN 759
Db    674 YYPSSTISLSDDFPHTI-AGEPDSRS-------SSSIASTVEIS-----SDLVSLTSDP- 719

Qy    760 PTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS----- 814
Db    720 ----TSSFDSSSSLNSDSSSSP-----------FSDESDISASSSFSTLVAPSFSLSSSS 764

Qy    815 -----RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKE 869
Db    765 SLSLIYPHYVNSTTYHASESESSSVASP-SVASESANDDTHTL-SESTDTTSIIGTDSST 822

Qy    870 ---CSRLSSSSGIVT-----NPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNV-- 919
Db    823 VTFCRRDNGDGCIVTGITSSSIDSEQTSDVTT---TSSFVASSTPTSAEQSITDNPNIDS 879

Qy    920 SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD----NQS 975
Db    880 SQTSASSSTKSSVSVSDTVVNSILLSETS--TLSSDDGTS-SDTSISSTTNSDTGNINAG 936
```

```
Qy    976 GTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSEPTNTR------EQPTTLSTTS 1026
Db    937 SSHKSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELTTIEDNEPNTFTSTP 996

Qy   1027 NSITEDITTSQPTGDNG------DNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
Db    997 SSHSEIFSS-----DNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTE------ASTA- 1044

Qy   1081 KPSMGENS-------GLTTSTEIEATTTSPTEAPSPAVSSGT--------DVTTEPTDT- 1124
Db    1045 --TLGENSFSNVASTPLNTATSLRSTSSSSNHATE---SSGTVKSEASVEAIPSPPTSTD 1099

Qy   1125 --------REQPTTLSTTSKTN-----SELVATTQATN---ENGGKSPSTDLTSSLTTGT 1168
Db    1100 NRLSYSTEEAEGITYANSGSTNNLITESQVAAPTDSTSVLIENLVVTSTFDDNSSAAVDQ 1159

Qy   1169 SASTSANSELV----TSGSVTGGAVASASNDQSHSTSV----------------TNSNSI 1208
Db    1160 PSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSA 1219

Qy   1209 VSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
Db    1220 ASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKHPSWLLKFISVALFF 1268


RESULT   5
Q9HGK6
ID   Q9HGK6      PRELIMINARY;      PRT;  2297 AA.
AC   Q9HGK6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALS7P.
GN   ALS7.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF201684; AAF98068.1; -.
SQ   SEQUENCE   2297 AA;  244723 MW;  59B020C63027F651 CRC64;


  Query Match             39.7%;  Score 2576.5;  DB 3;  Length 2297;
  Best Local Similarity   41.3%;  Pred. No. 8.7e-111;
  Matches  595;  Conservative  220;  Mismatches  416;  Indels  211;  Gaps   24;

Qy      2 LQQFTLLFLYLSIAS--AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGT 59
Db      1 MKKLYLLLYLLASFTTVISKEVTGVFNQFNSLIWSYTYRARYEEISTLTANAQLEWALDGT 60

Qy     60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 119
Db     61 IASPGDTFTLVMPCVYKFMTYETSVQLTANSIAYATCDFDAGEDTKSFSSLKCTVTDELT 120

Qy    120 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
Db    121 EDTSVFGSVILPIAFNVGGSGSKSTITDSKCFSSGYNTVTFFDGNNQLSTTANFLPRREL 180

Qy    180 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 239
Db    181 AFGLVVSQRLSMSLDTMTNFVMSTPCFMGYQSGKLGFTSNDDDFEIDCSSIHVGITNEIN 240

Qy    240 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 299
Db    241 DWSMPVSSVPFDHTIRCTSRALYIEFKTIPAGYRPFVDAIVQIPTTEPFFVKYTNEFACV 300

Qy    300 GSRLQSKPFTLRWT-GYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 358
Db    301 NGIYTSIPFTSFFSQPILYDEALAIGADLVRTTSTVIGSITRTTTLPFISRLQKTKTILV 360

Qy    359 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRT 418
Db    361 LEPIPTTTVTTSHHGFDTWYYTKKATIGDTATVFIDVPQHTATTLTTYWQESSTATTTYF 420

Qy    419 NPTDSIDTVVVQVPLP----------------------------------NPTVSTTE 442
Db    421 DDIDLVDTVIVKIPYPNPTIITTQFWSGKYLTTETHKEPPLGTDSVIIKEPHNPTVTTTE 480

Qy    443 YWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVII 502
Db    481 FWSESFATTETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIV 540

Qy    503 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTV 562
Db    541 REPHNPTVTTTEFWSESFATTETITNGPEGTDSVIVREPHNPTVTTTEFWSESFATTETI 600

Qy    563 TAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTE 622
Db    601 TNGPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIIKEPHNPTVTTTK 660

Qy    623 YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI 682
Db    661 FWSESFATTETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIV 720

Qy    683 REPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTV 742
Db    721 REPHNPTVTTTEFWSESFATTETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETV 780

Qy    743 TAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIY----ESMSSSK 798
Db    781 TNYPEGTDSVIVREPHNPTVTTTEFWSESFVTTETITTGPLGTDSIVIHDPLEESSSSTA 840

Qy    799 ISTSSNDITSI----------------------------IPS-------------FSRP 816
Db    841 IESSDSNISSSAQESSSSVEQSFTSADETSSIVELSSRSDIPSSSIGLTSSESSTVSSYD 900

Qy    817 HYVNSTT-----SDLSTFESSSMNTPTSISSD--GMLLSSTT------------------ 851
Db    901 SYSSSTSESSIASSYDSYSSSSIESSTLSSSDRYSSSISDTTSFWDSSSSDLESTSITWS 960

Qy    852 ------------LVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAS 899
Db    961 SSIDAQSSHLVQSVSNSISTSQEISSSSSEEESSTSATDALVSSDASSILSSDTSSYYPSS 1020

Qy    900 TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQL----SSI 955
Db   1021 TISPSDDFPHTI-AGESDSQSISFITSTVEISSDSVSLTSDP-ESSFDSSSRLNSDSSSS 1078

Qy    956 PS------VSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP--------- 1000
Db   1079 PSTDQRDILTSSSFSTLIKSSGSRESSIGTILSEESSDSIPTTFSTRYWSPSGMSSRHYT 1138

Qy   1001 ---------VVSSNT--DLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSST 1049
Db   1139 NSTETSVSDVVSSSVAGDETSESSVSVISESSESVTSESVASESVASESVASESVASESV 1198

Qy   1050 NPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSP 1109
Db   1199 TAVSDI--SDLYTTSEEVSTSDSNSGMSSPI-PSSEQRS------SIPIMSSSDESSESR 1249

Qy   1110 AVSSGTDVTTEPTDTREQPTTLST----------------TSKTNSELVATTQATNENGG 1153
Db   1250 ESSSGTILSEENSDS--IPTTFSTRYWSPSGMSSRHYTNSTETSVSDVVSSSVAGDETSE 1307

Qy   1154 KSPS--TDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNS---- 1207
Db   1308 SSVSVISESSESVTSESVASESVASESVASESVTAVSDISDLYTTSEVVSTSDSNSGMSS 1367

Qy   1208 ------------IVSNTPQ----------TTLSQQVTSSSPSTNTFIASTYDGSGSIIQH 1245
```

```
                    ::|::: :          | ||:: : | |:  ||       ||   :|
Db   1368 PIPSSEQRSSIPVMSSSDESSESRESSSGTILSEENSDSIPT--TFSTRYLSPSGMSSRH 1425

Qy   1246 ST 1247
            |
Db   1426 YT 1427


RESULT   6
Q9URQ0
ID   Q9URQ0      PRELIMINARY;      PRT;   468 AA.
AC   Q9URQ0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
DR   EMBL; AF024582; AAC64237.1; -.
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50143 MW;  2E38E36D7FEAD192 CRC64;


  Query Match             30.6%;  Score 1986.5;  DB 3;  Length 468;
  Best Local Similarity   80.0%;  Pred. No. 2.4e-84;
  Matches  375;  Conservative   33;  Mismatches   60;  Indels    1;  Gaps    1;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           || || || | :|:|:|| |||||:||:|||| | |||:||| ||||||||||||||
Db       1 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           ||||||||||||||| | |||||||:|||||||||||||||||||:| |||:: | |
Db      61 ANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           |||| ||||||:|||||||| ||| |:|| |||||||||||| |||||:|||:  |
       121 SIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 180

       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           | :  |||::||:|||:  :|||| |||||| ||| ||  |||||:|:||||||||
Db     181 SGFFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||:||||:| || |||:| || | |:|||||||| | | : |:  | ||||| | |
Db     241 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVG 299

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
           | |||||| ||  || ||:| ||| ||||||||||||||||||||||||||||||||||||
Db     300 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 359

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||||| ||||||||||||||||||||||||||||||| |||||||
Db     360 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITATTTRTNP 419

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
           |||||||||||| ||||||||||||||:|||||||||||||||:||||||
Db     420 TDSIDTVVVQVPSPNPTVSTTEYWSQSYATTTTVTAPPGGTDSVIIREP 468


RESULT   7
Q9URP8
```

```
ID   Q9URP8      PRELIMINARY;      PRT;   469 AA.
AC   Q9URP8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
DR   EMBL; AF024586; AAC64241.1; -.
FT   NON_TER     469    469
SQ   SEQUENCE   469 AA;  49597 MW;  88BC96D79142C8DB CRC64;


  Query Match             28.1%;  Score 1823;  DB 3;  Length 469;
  Best Local Similarity   72.1%;  Pred. No. 8e-77;
  Matches  338;  Conservative   49;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           || || || | :|:|:|| |||||:|||||||:|||:| ::||  ||| ||:|||||| :
Db       1 MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPATPTWTAVIGWSLDGAT 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           |: ||||||:||||||: | |||:|| |||  ||||   | |||||||:::|||   : :
Db      61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
             || |||||| :|:|||:||  || :|:||||| |||||||||  ||  |:||||||
Db     121 DTKAMGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVAS 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           |  :  ||::|||::  :||:  :| ||||||||||::     |||| :|:||: ||||
Db     181 SDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||||:|||||||||||| : : ||||||||||:|||:||| |: ||: ||| | | |
Db     241 WNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFVDAYVSATRVSSYTMQYTNIYACVG 300

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
           :       ||  | || || |||||| || |||||||||||||||||   |||||||||
Db     301 AASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSDTDKTKTIEILQ 360

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           |||||||||||||||||| ||||||||||||||||||||||||||||||| |||||||
Db     361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
           |||||||||||| ||||||||||||||:|||||||||||||||:||||||
Db     421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREP 469


RESULT   8
Q9C471
ID   Q9C471      PRELIMINARY;      PRT;   468 AA.
AC   Q9C471;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS9.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
```

```
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RA   Hoyer L.L., Hecht J.E., Mirus K.A.;
RT   "The ALS9 gene of Candida albicans.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229989; AAK00764.1; -.
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50127 MW;  B291D3EB15FB96DE CRC64;

  Query Match           27.4%;  Score 1780.5;  DB 3;  Length 468;
  Best Local Similarity   70.1%;  Pred. No. 7.2e-75;
  Matches  329;  Conservative   62;  Mismatches   77;  Indels    1;  Gaps    1;

Qy       1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
           || || |||: |::::|||||||:||:||||: :  ||:|||  ||||||||||: |:
Db       1 MLPQFILLFISLTVSTAKTITGVFNSFDSLTWTRSVEYAYKGPETPTWNAVLGWSLNSTT 60

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           |:|||||| :|||||: |:|||||||||| |||| | :|||||||:|:|||||    |
Db      61 ADPGDTFTLILPCVFKFITTQTSVDLTADGVSYATCDFNAGEEFTTFSSLSCTVNSVSVS 120

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
             :  ||| ||| |||||||||| || |||||||| ||||| |||  ||  |:|: | | |
Db     121 YARVSGTVKLPITFNVGGTGSSVDLADSKCFTAGKNTVTFMDGDTKISTTVDFDASPVSP 180

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           | |: :||::|||||:::||| |||||||||| ||| :|||   |||||::|||:|||||
Db     181 SGYITSSRIIPSLNKLSSLFVVPQCENGYTSGIMGFVASNG-ATIDCSNVNIGISKGLND 239

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||:|||||||||||||  | :::|||||||||||:|||||| :::::||| | |:|||
Db     240 WNFPVSSESFSYTKTCTSTSITVEFQNVPAGYRPFVDAYISAENIDKYTLTYANEYTCEN 299

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
                 |||| | |||||:| |:| ||| |||||||||||||||||||||||:||||||
Db     300 GNTVVDPFTLTWWGYKNSEADSDGDVIVVTTRTVTDSTTAVTTLPFNPSVDKTETIEILQ 359

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
           ||||||||||:|::||| | |  || ||||||| ||| | |||: |||:|||||| |||
Db     360 PIPTTTITTSYIGISTSYETLTGTIGGTATVIVDTPYHITATVTNFWTGSITTTTTYTNP 419

Qy     421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
           | |||||:||:| |:|| : ||:||:|||:|||:| || ||::|||:||
Db     420 TGSIDTVIVQIPSPDPTTTITEFWSESFASTTTITNPPDGTNSVIIKEP 468


RESULT  9
Q9HF70
ID   Q9HF70      PRELIMINARY;      PRT;   336 AA.
AC   Q9HF70;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALSD2P (FRAGMENT).
GN   ALSD2.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202529; AAG35623.2; -.
FT   NON_TER       1      1
FT   NON_TER     336    336
SQ   SEQUENCE   336 AA;  35593 MW;  E513E6EA9E8E9EC7 CRC64;

  Query Match           22.6%;  Score 1471;  DB 3;  Length 336;
  Best Local Similarity   81.0%;  Pred. No. 9.1e-61;
  Matches  272;  Conservative   31;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           |:|||||| |||||: |:|||||||:|||||| |||||||||||:|||||| || |
Db       1 ASPGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCSFYSGEEFTTFSSLTCTVNSALTS 60

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           |:|||||||||:|||||||| :|:|||||||||:||| |||  :|  |:|||||||
Db      61 SVKAFGTVTLPISFNVGGTGSSVDLEDSKCFTAGTNTVTFTDGDNKVSTTVDFEKSTVDS 120

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           : || :||:|||||||:|||||||  ||||||:|||||| |: ||||:|:||| |:||
Db     121 TGYLTSSRLMPSLNKVTSLFVAPQCARGYTSGTIGFSSSNGGVSFDCSNVHVGITNGVND 180

Qy     241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
           ||:|||||||||||:| || | |:|||||||||||:|:||| |||:|||||| |
Db     181 WNFPVSSESFSYTKTCSSTGITITYRNVPAGYRPFIDAYITASDVNSYTLSYTNDYTCVG 240

Qy     301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
             :| |||||||:|||| :|||||||||||||||||||||||  ||||||||||
Db     241 GSVQHKPFTLRWSGYKNGEAGSNGIVIVATTRTVTDSTTAVTTLPFNSDVDKTKTIEILQ 300

Qy     361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 396
           |||||||| |||||||| |||||||:|||: :|:|
Db     301 PIPTTTITTLYVGVTTSYSTKTAPIGDTATLFIDIP 336


RESULT  10
Q9HF72
ID   Q9HF72      PRELIMINARY;      PRT;   338 AA.
AC   Q9HF72;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALSD1P (FRAGMENT).
GN   ALSD1.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201685; AAG35603.1; -.
FT   NON_TER       1      1
FT   NON_TER     338    338
SQ   SEQUENCE   338 AA;  35908 MW;  C0A7F2F94609E172 CRC64;

  Query Match           18.3%;  Score 1191;  DB 3;  Length 338;
  Best Local Similarity   64.8%;  Pred. No. 7.1e-48;
  Matches  219;  Conservative   51;  Mismatches   66;  Indels    2;  Gaps    1;

Qy      61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
           || |||||| |||||: |:|||||||:|||||| |::||:|| ||:::| ||: | |
Db       1 ANAGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCTFHAGEDFTAFSSMSCVVNNGLTS 60

Qy     121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
           :|||||| :||:|||||||| :|:|||||||||:||| |||  ||| |:| |:
Db      61 NIKAFGTVRIPISFNVGGTGSSVNLQDSKCFTAGTNSVTFTDGDHKISIPVDFPKTPESS 120

Qy     181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
           |  :  |||:|:|:|:::| || ||  || || :|||::  || |:|||:|:||| |||
Db     121 SGLIKYSRVIPTLDKLSSLAVASQCTAGYKSGVLGFSATKNDVTIECSNVHVGITNGLNS 180
```

```
Qy    241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQ--YTLAYTNDYTC 298
          || ||||:|||||||||||:    | |:|||||||||||| |:  |      : | ||| | |
Db    181 WNMPVSSDSFSYTKTCTSSSFIITYENVPAGYRPFIDTYVKKTSTTSTGFNLNYTNSYVC 240

Qy    299 AGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 358
            :  : |    || | ||||||:| |:: ||:|||||||:|||||:|:|||||||:
Db    241 TDGKKGNDPLIYFWTSYTNSDAGSDGAVVIVTTKTVTDSTTAITTLPFDPTVDKTKTIEV 300

Qy    359 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 396
          |:|||||||||||||:||: |||| ||||||:|:|||
Db    301 LEPIPTTTITTSYVGVSTSPFTKTATIGETATLIIDVP 338


RESULT 11
Q9HF69
ID   Q9HF69      PRELIMINARY;      PRT;   331 AA.
AC   Q9HF69;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALSD3P (FRAGMENT).
GN   ALSD3.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202530; AAG35624.2; -.
FT   NON_TER       1      1
FT   NON_TER     331    331
SQ   SEQUENCE   331 AA;  35297 MW;  469AB72F9CE029BF CRC64;


  Query Match             18.1%;  Score 1176;  DB 3;  Length 331;
  Best Local Similarity   64.3%;  Pred. No. 3.4e-47;
  Matches  214;  Conservative   44;  Mismatches   73;  Indels    2;  Gaps    1;

Qy     61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
          || |||||||||||||: | |||||| |||  ||||  ||||||||||:| |||::|| |
Db      1 ANAGDTFTLNMPCVFKFITDQTSVDLVADGRTYATCDLYSGEEFTTFSSLKCTVSNALNS 60

Qy    121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
            || |||||::||:||:||  |:  |:||  |||||||||||   |    |::| |:
Db     61 QTKALGTVTLPLSFNIGGSGSDVDITSSQCFKEGTNTVTFNDGDTTFSTTANFQRSDVNA 120

Qy    181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
          :  :  ||::||| |  |:|: |:| :||:|||||||::  |  |||| :| ||: ||||
Db    121 NDRILLSRILPSLAKSVTIFIPPRCASGYSSGTMGFSTAGTDAIIDCSTVHAGISNGLND 180

Qy    241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
          |||| |:|||||| :|:| |: : |||||||||||:||||||  :  ||: ||| ||| |
Db    181 WNYPIESKSFSYTTSCSSKGVSVTYQNVPAGYRPFVDAYISA--LTSYTMQYTNQYTCVG 238

Qy    301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
          :|     |:  | || |::||| || || || |||||||||||||||  ||:|||| :||
Db    239 ARPVDASFSYNWLGYDNAEAGSRGITIVVTTSTVTDSTTAVTTLPFNSDVDRTKTIAVLQ 298

Qy    361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIV 393
          |||||||||||||||| |:| |||||||: :
Db    299 PIPTTTITTSYVGVTTSYRTQTVPIGETATLFI 331


RESULT 12
Q9Y8F1
ID   Q9Y8F1      PRELIMINARY;      PRT;   433 AA.
AC   Q9Y8F1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN 7 (FRAGMENT).
GN   ALS7.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF075294; AAD42034.1; -.
FT   NON_TER     433    433
SQ   SEQUENCE   433 AA;  48037 MW;  1F30CFD99C2EC445 CRC64;


  Query Match             15.4%;  Score 999.5;  DB 3;  Length 433;
  Best Local Similarity   46.0%;  Pred. No. 6e-39;
  Matches  199;  Conservative   76;  Mismatches  155;  Indels    3;  Gaps    2;

Qy      2 LQQFTLLFLYLSIAS--AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGT 59
            :::  ||:|  |  :  :| :||||: |||| ||       ::      | || | |:||||
Db      1 MKKLYLLLYLLASFTTVISKEVTGVFNQFNSLIWSYTYRARYEEISTLTANAQLEWALDGT 60

Qy     60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 119
           |:|||||| ||||:|: | :||| |||: : |||| | :||:  :||:| ||| | |
Db     61 IASPGDTFTLVMPCVYKFMTYETSVQLTANSIAYATCDFDAGEDTKSFSSLKCTVTDELT 120

Qy    120 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
              ||:| ||||||||||:|| : : |||||::| ||||| ||:  :|    |
Db    121 EDTSVFGSVILPIAFNVGGSGSKSTITDSKCFSSGYNTVTFFDGNNQLSTTANFLPRREL 180

Qy    180 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 239
              : : |:  ||: :|   ::  |  || || :||:|:: |  ||||:||:|||  :|
Db    181 AFGLVVSQRLSMSLDTMTNFVMSTPCFMGYQSGKLGFTSNDDDFEIDCSSIHVGITNEIN 240

Qy    240 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 299
          ||: ||||  | :|  |||  : |::: :||||||||:|| :       : : |||:: |
Db    241 DWSMPVSSVPFDHTIRCTSRALYIEFKTIPAGYRPFVDAIVQIPTTEPFFVKYTNEFACV 300

Qy    300 GSRLQSKPFTLRWT-GYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 358
              | |||  ::     :| : |  :| || ||  | |  |||||   : ||||| :
Db    301 NGIYTSIPFTSFFSQPILYDEALAIGADLVRTTSTVIGSITRTTTLPFISRLQKTKTILV 360

Qy    359 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRT 418
          |:|||||:|||: |  | | || | ||:|||| :||| || ||:|: |  : | |||
Db    361 LEPIPTTTVTTSHHGFDTWYYTKKATIGDTATVFIDVPQHTATTLTTYWQESSTATTTYF 420

Qy    419 NPTDSIDTVVVQV 431
          :  | :|||:|::
Db    421 DDIDLVDTVIVKI 433


RESULT 13
Q9HF71
ID   Q9HF71      PRELIMINARY;      PRT;  353 AA.
AC   Q9HF71;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN 1 (FRAGMENT).
GN   ALST1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13803;
```

```
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201686; AAG35604.1; -.
FT   NON_TER      1      1
FT   NON_TER    353    353
SQ   SEQUENCE   353 AA;  37272 MW;  4268154D5615DE08 CRC64;

  Query Match             15.1%;  Score 981;  DB 3;  Length 353;
  Best Local Similarity   54.9%;  Pred. No. 3.5e-38;
  Matches  185;  Conservative   50;  Mismatches  100;  Indels    2;  Gaps    2;

Qy        60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 119
             :|| |||||| |||||||:|||:||:|||       |||| | :|| ||||||:|:||| ::
Db        19 AANAGDTFTLIMPCVFKFTTSETSIDLTVGSKSYATCNFNAGEHFTTFSSLSCTVTQSVP 78

Qy       120 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
              : |:||:|:|:||||||:|   || |:||||  | ||||||:||||  |     || :
Db        79 DNTNAYGTITVPLAFNVGGSGRDVDLTDAKCFTTGDNTVTFSDGDKSFSTTANFEGAGTL 138

Qy       180 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 239
               | :||::|||  |  | |||  |  |||  |||:||||:    :|||:||  |||  ||
Db       139 NDDY-ESSRLIPSLGKTDALLVAPLCSNGYKSGTIGFSSTTKGFSIDCNNIQAGITSQLN 197

Qy       240 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 299
              | :|  |:|||||  ||:     | :   :| |  ||||||||  |    :  | :  ||  | |:
Db       198 AWGFPTDSQSFSYTTQCTTTSYSITFSTIPKGLRPFIDAYIKA-PTSTYPMTYTYKYVCS 256

Qy       300 GSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEIL 359
              :  :      | |:|| |||| | |:  ||   | |  | |||  ||||||  :  ||||||:::
Db       257 DGKSYNGNTKLNWSGYVNSDADSEGMEIVVATTGTGSTTGVTTLPFDKTKDKTKTIQVI 316

Qy       360 QPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 396
             :||||||:||||:|||||:  | || ||||||:::|:|
Db       317 EPIPTTTVTTSYLGVTTSFSTITATIGETATLVIDMP 353

RESULT 14
Q96WU8
ID   Q96WU8      PRELIMINARY;      PRT;  1195 AA.
AC   Q96WU8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 119.8 KDA PROTEIN (FRAGMENT).
GN   SPCPB16A4.07C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL591677; CAC39326.1; -.
KW   Hypothetical protein.
FT   NON_TER    1195   1195
SQ   SEQUENCE   1195 AA;  119807 MW;  5D8892EAA086E835 CRC64;

  Query Match             12.8%;  Score 831.5;  DB 3;  Length 1195;
  Best Local Similarity   28.7%;  Pred. No. 9.6e-31;
  Matches  319;  Conservative  154;  Mismatches  446;  Indels  193;  Gaps   48;

Qy       251 SYTKTCT---SNGIQIKYQNVPA----GYRPFIDAYISATDVNQYTLAYTNDYTCAGSRL 303
              | | | |    |||  |    ||         | :|  :   |: :: : |    :
Db         1 SVTTTLTGAPSNGTVIDTVEVPTTTNYGYTTVTTGYTGTSTTLTT-TVPHSGNETGPTTVY 59

Qy       304 QSKPF-----TLRWTGYKNS------DAGSNGIVI------------VATTRTVTDSTT 339
               |:     |    ||  |||     | || ||                |  | |||
Db        60 VETPYPTTVTTTTVGYPGSVTTTLTGAPSNGTVIDTVEIPTTTNYGYTTITTGYTGSTT 119

Qy       340 AVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI--VDVPY 397
              ||:| : :        |: :  | |||  ||: ||  |  |           |||  |:||
Db       120 LTTTVPHSGNETGPTTVYVETPYPTTVTTTTVGYPGSVTTTLTGAPSNGTVIDTVEVPT 179

Qy       398 HTT---TTVTSEWTGTITTTTT------RTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSF 448
              |     ||:|: :||: | |||         | ||     || |: | |      :||
Db       180 TTNYGYTTITTGYTGSTTLTTTVPHSGNETGPT----TVYVETPYPTTVTTTTVGYPGS 235

Qy       449 ATTTTVTAPPGGT--DTVIIREPPN--HTVTTTEYWSQSFATTTTVTAPPGGTD----SV 500
             |||   || ||  |||  :       |  :|  || |    : :|| | | |  | :   :|
Db       236 VTTTLTGAPSNGTVIDTVEVPTTTNYGYTTVTTGY---TGSTTLTTTVPHSGNETGPTTV 292

Qy       501 IIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------W 552
              :  |   |||||   | : |||:|   |::|::        || |: |            |
Db       293 YVETPYPTVTTTTVGYSGSVTTTLTG--SGSNSIVTETVDVPTTTSVNYGYTTITTGW 350

Qy       553 SQSYATTTTVT--APPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVII 610
              :  |: ||    | :| |  |        |||      | : |||:|   |   |  :
Db       351 TGSTTLTSIVTHSGSETGPTTVYIETPSVSATTTTTIGYSGSLTTTLTGSSGPVVTNTV 410

Qy       611 REPPNH--------------TVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPT--- 653
              |  :              |  | ||  |    :| | | :   :|| ||:|:|  |   |
Db       411 EIPYGNSSYIIPTTIVTGTVTTVTTGYTGTETSTVTVIPTGTTGTTTVVIQTPTTVTATE 470

Qy       654 ---VTTTEYWSQSYATTTTITAPPGE---TDTVLIREPPNHT------VTTTEYWSQSYA 701
               || |   :: :   :| |:| ||   |  | ||:|: |        || |     :: :
Db       471 TDIVTVTTGYTGTETSTVTVT-PTGTSTGTTTVVIQTPTTVTATETDIVTVTTGYTGTET 529

Qy       702 TTTTVTAPPGE---TDTVLIREPPNHT------VTTTEYWSQSYATTTTVTAPPG---GT 749
             :| ||| | |    |  | ||:|:  |    |       ||  ::  :  :| ||| | |   ||
Db       530 STVTVT-PTGTSTGTTTVVIQTPTTVTATETDIVTVTTGYTGTETSTVTVT-PTGTSTGT 587

Qy       750 DTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSI 809
              ||:|: |    | | |:             |||   |||:|      ::: :   ||:   | :
Db       588 TTVVIQTPTTVTATETD--------IVTVTTGYTGTET----STVTVTPTGTSTGTTTVV 635

Qy       810 IPSFSRPHYVNSTTSDLSTFESSSMNTPTS---ISSDGMLLSSTTLVTESETT-----TE 861
              | :  |  :| :|: | :   :  |   |  ||   :: :   :|||:| :: ||    |:
Db       636 IQT---PTTVTATETDIVTVTTGYTGTETSTVTVTPTGTSTGTTTVVIQTPTTVTATETD 692

Qy       862 LICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAS---TMSDSLSSTDGISATSSDN 918
              ::           :  |    |:| || | ||  :   |  :  :  :  ::|| :|
Db       693 IVT-------------VTTGYTGTETSTVTVT-PTGTSTGTTTVVIQTPTTVTATETDI 737

Qy       919 VSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTH 978
              |: :    ||||  |:    |  ::    :  | : :   |: | ||  |  |:    :||
Db       738 VTVTTGYTGTETSTVTV----TPTGTATGTTTVVINTPTTGSE--VLPTTGATGTAGTE 791

Qy       979 DSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLS-TTSNSITEDITTSQ 1037
              :|:||:: || :|         :  | :|    |  |  | ||| | : : :  || ||:
Db       792 TQLTTATEVQ-PTTGAT----GTAGTETQVT---TGTETQATTATETQATTATEVQTTTG 843

Qy      1038 PTGDNGDNTSSTNPV---PTV-ATST------LASASEEDNKSGSHESASTSLKPSMGEN 1087
              ||  |  | :|     || | ||  |||     ||   :: : :|  :|   :| | : : |
Db       844 ATGTAGTETQATTATEVQPTTGATGTAGTETQVTTATEVQPTTGATGTAGTETQVTTGTE 903

Qy      1088 SGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLST-TSKTNSELVATTQ 1146
              : ||:||  :||| :   :  : |  :    |    | |  ||||  :| |: |  ::: |:
Db       904 TQATTATETQATTATEVQTTTGATGTAGTETQATTATEVQPTTGATGTAGTETQVTTATE 963

Qy      1147 ---ATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVT 1203
                |  |      | :|:  | : :|  |   |  |||  :|  :    |||
Db       964 VQPTTGATGTAGTETQVTTGTETQATTATETQATTATEVQTTTGATGTAGTE----TQVT 1019

Qy      1204 NSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
              :  :      | | :   ||||   | ::||
Db      1020 TATEV-----QPTTAVTETSSSGYYTTIVSST 1046

RESULT 15
```

```
Q9HG16
ID   Q9HG16      PRELIMINARY;      PRT;   226 AA.
AC   Q9HG16;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1177;
RA   Hoyer L.L., Kapteyn J.C., Hecht J.E., En J., Klis F.M.;
RT   "Evidence Suggesting the Presence of an ALS Gene Family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF189016; AAG17111.1; -.
FT   NON_TER       1      1
FT   NON_TER     226    226
SQ   SEQUENCE   226 AA;  24802 MW;  2AA36A42E44CB460 CRC64;


  Query Match             12.7%;  Score 824;  DB 3;  Length 226;
  Best Local Similarity   70.1%;  Pred. No. 3.7e-31;
  Matches  155;  Conservative   28;  Mismatches   38;  Indels    0;  Gaps    0;

Qy     429 VQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTT 488
           |||| ||||||:||::|| |  || |||  | |||:|||:|| | |||||||:||:|||||
Db       1 VQVPSPNPTVTTTQFWSGSVPTTETVTTGPQGTDSVIIKEPHNPTVTTTEFWSESFATTE 60

Qy     489 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT 548
           |||  | |||||||:|| ||||||||:||:||||| |||  | ||||||:||| ||||||
Db      61 TVTNNPEGTDSVIIKEPHNPTVTTTEFWSESFATTETVTNYPEGTDSVIVREPHNPTVTT 120

Qy     549 TEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTV 608
           ||:||:|:||| |||  | ||||||:||| | ||||||:||:|:||| |:|  | |||:|
Db     121 TEFWSESFATTETVTNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNYPEGTDSV 180

Qy     609 IIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREP 649
           |:||| | ||||||:||:|| || |:|  | |||:::| :|
Db     181 IVREPHNPTVTTTEFWSESFVTTETITTGPLGTDSIVIHDP 221


Search completed: October  3, 2002, 15:37:25
Job time: 306 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:25:37 ; Search time 31.64 Seconds
(without alignments)
1460.389 Million cell updates/sec

Title: US-09-715-876-8_COPY_17_432
Perfect score: 2190
Sequence: 1 AKTITGVFDSFNSLTWSNAA.....TTTTRTNPTDSIDTVVVQVP 416

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	308.5	14.1	650	15	AAR47575	Alpha-agglutinin o
2	177	8.1	1537	15	AAR60562	Yeast 4.7 kB agglu
3	175.5	8.0	1837	21	AAB11726	Cryptosporidium pa
4	175	8.0	1721	19	AAW48299	Cryptosporidium pa
5	175	8.0	1721	21	AAB11727	Portion of Cryptos
6	156	7.1	688	22	ABB30137	Peptide #2788 enco
7	156	7.1	688	22	ABB35307	Peptide #2813 enco
8	156	7.1	688	22	ABB20749	Protein #2748 enco
9	156	7.1	688	22	AAM56138	Human brain expres
10	156	7.1	688	22	AAM68511	Human bone marrow
11	156	7.1	688	22	AAM16315	Peptide #2749 enco
12	156	7.1	688	22	AAM28810	Peptide #2847 enco
13	156	7.1	688	22	AAM04053	Peptide #2735 enco
14	155	7.1	894	15	AAR47578	Flocculation prote
15	155	7.1	894	15	AAR58754	S. cerevisiae FLO1
16	153.5	7.0	1795	22	ABB69806	Drosophila melanog
17	151	6.9	2586	22	ABB66878	Drosophila melanog
18	150.5	6.9	175	21	AAB11743	C. parvum Iowa iso
19	150.5	6.9	288	20	AAY29081	T. gondii immunoge
20	150.5	6.9	288	20	AAY29082	T. gondii immunoge
21	150.5	6.9	288	22	AAU25552	T. gondii immunoge
22	150.5	6.9	288	22	AAU25553	T. gondii immunoge
23	150	6.8	386	22	ABB29832	Peptide #2483 enco
24	150	6.8	386	22	ABB35008	Peptide #2514 enco
25	150	6.8	386	22	ABB20422	Protein #2421 enco
26	150	6.8	386	22	AAM55821	Human brain expres
27	150	6.8	386	22	AAM68195	Human bone marrow
28	150	6.8	386	22	AAM16015	Peptide #2449 enco
29	150	6.8	386	22	AAM28516	Peptide #2553 enco
30	150	6.8	386	22	AAM03749	Peptide #2431 enco
31	148.5	6.8	216	21	AAB11729	Cryptosporidium pa
32	148.5	6.8	862	15	AAR60563	Yeast 2.6 kB agglu
33	146.5	6.7	249	21	AAB11746	C. parvum NINC iso
34	146.5	6.7	1045	22	ABG08332	Novel human diagno
35	146.5	6.7	1532	21	AAB40945	Human ORFX ORF709
36	145	6.6	130	21	AAB11740	C. parvum Iowa iso
37	145	6.6	1959	12	AAR10562	Mutant protease (d
38	145	6.6	1962	12	AAR10560	Mutant protease (K
39	145	6.6	1962	12	AAR10561	Mutant protease (N
40	145	6.6	1962	12	AAR10557	Mutant protease (A
41	145	6.6	1962	12	AAR10558	Mutant protease (A
42	145	6.6	1962	12	AAR10559	Mutant protease (A
43	145	6.6	1962	12	AAR10563	Mutant protease (K
44	145	6.6	1968	12	AAR10941	Mutant protease (d
45	145	6.6	1974	12	AAR10940	Mutant protease (d

ALIGNMENTS

RESULT 1
AAR47575
ID AAR47575 standard; Protein; 650 AA.
XX
AC AAR47575;
XX
DT 19-JUL-1994 (first entry)
XX
DE Alpha-agglutinin of Saccharomyces cerevisiae.
XX
KW Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW inulinase; alpha-amylase; Saccharomyces cerevisiae;
KW enzymatic process; fermentation; biodegradation; catalysis.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9401567-A.
XX
PD 20-JAN-1994.
XX
PF 07-JUL-1993; 93WO-EP01763.
XX
PR 08-JUL-1992; 92EP-0202080.
PR 14-DEC-1992; 92EP-0203899.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
PI Klis FM, Schreuder MP, Toschka H, Verrips CT;
XX
DR WPI; 1994-035071/04.
DR N-PSDB; AAQ54012.

XX
PT Immobilisation of enzymes to microbial cell wall - by prodn. of
PT fusion protein of enzyme linked to anchoring protein
XX
PS Example 1; Page 32-39; 99pp; English.
XX
CC The alpha-agglutinin is used in a method to immobilise enzymes to a
CC microbial cell wall. The coding sequence is used in the production
CC of a recombinant polynucleotide which comprises a structural gene
CC encoding a protein with catalytic activity and at least part of a
CC gene encoding at least the C-terminus of a protein capable of
CC anchoring in a eukaryotic or prokaryotic cell wall. The anchoring
CC fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1,
CC major cell wall protein of lower eukaryotes or a proteinase of
CC lactic acid bacteria. The recombinant polynucleotide preferably
CC also comprises a sequence encoding a signal peptide to ensure
CC secretion of the expressed product. The signal peptide is
CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC alpha-amylase of Bacillus or proteinases of lactic acid bacteria.
CC The host microorganism can be used for performing enzymatic
CC processes on an industrial scale.
XX
SQ Sequence 650 AA;

Query Match 14.1%; Score 308.5; DB 15; Length 650;
Best Local Similarity 25.8%; Pred. No. 2.7e-16;
Matches 116; Conservative 79; Mismatches 190; Indels 65; Gaps 20;

```
Qy      10 SFNSLTWSN------AANYAFKGPGYPTWNAVLGWSL-DGTSANPGDTFTLNMPCVFK-- 60
Db      21 ninditfsnleitpltan---kqpd-qgwtatfdfsiadassiregdeftlsmphvyrik 76

Qy      61 -YTTSQTSVDLTADGVKYATCQFYSGE-----EFTTFSTLTCTVNDALKSSIKAFGTVTL 114
Db      77 llnssqtatisladgteafkc-yvsqqaaylyenttf---tctaqndlssyntidgsitf 132

Qy     115 PIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY---LYAS 171
Db     133 slnfsdgggssyeyelenakffksgpmlvklgnqmsdv---vnf-----dpaaftenvfhs 184

Qy     172 RVMPSLNKVTTLFVAPQCENGYTSG---TMGFSSSNGDVAIDCSNIHIGITKGLNDWNYP 228
Db     185 grstgygsfesyhlgmycpngyflggtekidydssnnnvdldcssvqvyssndfndwwfp 244

Qy     229 VSSESFSYTKTCTSNGIQIKY-QNVPAGYRPFIDAYIS-ATDVNQYTLAYTNDYTC---- 282
Db     245 qsyndtnadvtcfgsnlwitldeklydgemlwvnalqslpanvntidhalefqytcldti 304

Qy     283 AGSRLQSKPFTLR----WTGYKNSDAGSNGIVIVATTRTVTD------STTAVTTLPFNP 332
Db     305 anttyatqfsttrefivyqgrnlgtasakssfistttdltsintsaystgsistv---- 360

Qy     333 SVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTT------- 385
Db     361 etgnrttsevishvvttstklsptattsltiaqtsiystdsnitvgtdihttsevisdve 420

Qy     386 TVTSEWTGTITTT-TTRTNPTDSIDTVVVQ 414
Db     421 tisretastvvaaptsttgwtgamntyipq 450
```

RESULT 2
AAR60562
ID AAR60562 standard; Protein; 1537 AA.
XX
AC AAR60562;
XX
DT 13-APR-1995 (first entry)
XX
DE Yeast 4.7 kB agglutination gene FLO1L.
XX
KW Yeast; agglutination; FLO1L.
XX
OS Saccharomyces cervisiae ABXL-1D.
XX
PN WO9419475-A.
XX
PD 01-SEP-1994.
XX
PF 24-FEB-1994; 94WO-JP00290.
XX
PR 26-FEB-1993; 93JP-0038871.
XX
PA (PANI-) PANIMOLABORATORIO BRYGGERILABORATORIUM.
PA (SAPB) SAPPORO BREWERIES.
XX
PI Keraenen S, Ogawa M, Onnela M, Penttila M, Takata Y;
PI Watari J;
XX
DR WPI; 1994-294338/36.
DR N-PSDB; AAQ71390.
XX
PT New yeast agglutination genes and yeast contg. them - impart
PT agglutination properties to facilitate removal from fermentation
PT media
XX
PS Disclosure; Page 43-48; 75pp; English.
XX
CC The agglutination gene is called FLO1. Saccharomyces cerevisiae
CC includes an agglutination gene of 4.7 kb (FLO1L) and an
CC agglutination gene of 2.6 kb (FLO1S). FLO1L is the intact FLO1 gene
CC on chromosome I, and FLO1S is the FLO1L gene with a portion of the
CC ORF deleted in frame. FLO1L imparts a relatively strong
CC agglutinative property to the host yeast into which it is introduced,
CC while FLO1S imparts a weaker agglutinative property.
XX
SQ Sequence 1537 AA;

Query Match 8.1%; Score 177; DB 15; Length 1537;
Best Local Similarity 23.0%; Pred. No. 3.5e-05;
Matches 112; Conservative 47; Mismatches 181; Indels 146; Gaps 25;

```
Qy      21 NYAFKGPG-YPTWNAVLGWSLD--GTSANPGDTFTLNM----------PCVFKYTTSQ-- 65
Db     104 nwgckgmgacsnsqgiaywstdlfgfyttptnv-tlemtgyflppqtgsytfkfatvdds 162

Qy      66 ---------------------TSVDLTADGVK----------YATCQFYSGEEFTTFSTL 94
Db     163 ailsvggatafnccaqqqppitstnftidgikpwggslppniegtvymyagyyypm---- 218

Qy      95 TCTVNDALKSSIKAFGT----VTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKD 150
Db     219 -----kvvysnavswgtlpisvtlp-----dgttvsddfeg--------yvysfdd---- 256

Qy     151 ISIDVEFEKSTV-DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI 209
Db     257 ---dlsqsnctvpdpsnya-vsttttttepwtgtftststemttvtgtngvptdetvivi 312

Qy     210 DCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDV 269
Db     313 rtpttastiitttepwnstftststelttvtgtngvrtdetii------virtpttatta 366

Qy     270 NQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVT 326
Db     367 itttepwnstftststel------------ttvtgtnglptdetiivirtpttattamt 413

Qy     327 TL-PFNPSVDKTKT----------------IEILQPIPTTTITT-------SYVGVTTS 361
Db     414 ttqpwndtftststemttvtgtnglptdetiivirtpttattamtttqpwddtftstste 473

Qy     362 YLTKTA----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTTTRTN--PTDS 407
Db     474 mttvtgtnglptdet-iivirtpttattamtttqpwndtftststemttvtgtnglptd- 531
```

```
Qy     408 IDTVVV 413
           :|:::|
Db     532 -etiiv 536


RESULT   3
AAB11726
ID   AAB11726 standard; Protein; 1837 AA.
XX
AC   AAB11726;
XX
DT   28-OCT-2000  (first entry)
XX
DE   Cryptosporidium parvum Iowa isolate GP900.
XX
KW   GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1785
FT                   /note= "Encoded by AAC in AAA61846"
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-0928361.
XX
PR   13-SEP-1996;   96US-0026062.
PR   01-JUN-1993;   93US-0071880.
PR   29-MAY-1992;   92US-0891301.
PR   03-APR-1995;   95US-0415751.
PR   14-AUG-1996;   96US-0700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   N-PSDB; AAA61846, AAA61847.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and
PT   diagnosing or treating Cryptosporidium infections by competitive
PT   inhibition of the function of GP900 -
XX
PS   Claim 2; Column 51-60; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents the GP900 protein of the Iowa isolate of Cryptosporidium
CC   parvum.
XX
SQ   Sequence   1837 AA;


  Query Match             8.0%;  Score 175.5;  DB 21;  Length 1837;
  Best Local Similarity   22.9%;  Pred. No. 5.9e-05;
  Matches   93;  Conservative   45;  Mismatches  176;  Indels   93;  Gaps   13;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
           |:  |  | |          ||: |:   |       |    : :: || :| | |
Db     452 ttkkpttttttttttttttttttttttttttttttttttattttkkpttttttttttttkk 511

Qy     103 KSSIKAFGTVTLPIAFNVGGTGSSTDL----------EDSKCFTAGTNTVTFNDGDKDI 151
           ::     | |         | | |:          :: :|   |    |      |  |
Db     512 pttttattttt--------ttsetesvikpdewcwleknqeceakgatyvgvigkdgri 562

Qy     152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI-- 209
           : |     | :   :  :|    | ::      :|  |  :| : |   : |  |
Db     563 engmaftmipnddthvrfrfkvkdvgntisv-----rcgkg--agklefpdrsldftipp 615

Qy     210 -----DCSNIHIGITKGLNDWNYPVSSESFS--------------YTKTCTSNGIQIKY 249
                || | :|:: |      |  |:  |               |  |||:     ||
Db     616 vaghnscs-iivgvsgggkihvspygskdvslisapiqpcelfnevycdtcta-----ky 669

Qy     250 QNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI 309
           : :||:   |   : |                 :|| |        :  |: |
Db     670 gaihsgyqtsadfvtttt-------------------akptt--------tttgapgq 700

Qy     310 VIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPI 369
           || : :  ||  ||         | |  || || ||| |      ||: :    ||
Db     701 pttttgspskpttttttt-------kattttilnpiittt-tqkpttttttkvpgkppi 752

Qy     370 GETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           | | :  :   |||  |:  | |: |||| |   | : |    :|
Db     753 atttttlkpivttttttkatttttttvptttt-ttkrdemttttplp 798


RESULT   4
AAW48299
ID   AAW48299 standard; Protein; 1721 AA.
XX
AC   AAW48299;
XX
DT   17-AUG-1998  (first entry)
XX
DE   Cryptosporidium parvum GP900 antigen.
XX
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; 3' region.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   Domain          175..423
FT                   /note= "NINC domain"
XX
PN   WO9806430-A1.
XX
PD   19-FEB-1998.
XX
PF   11-AUG-1997;   97WO-US14104.
XX
PR   14-AUG-1996;   96US-0700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Gut J,  Leech J,  Nelson RC,  Petersen C;
XX
DR   WPI; 1998-159290/14.
DR   N-PSDB; AAV20700, AAV20701.
XX
```

```
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
XX
PS   Claim 27; Pages 65-68; 89pp; English.
XX
CC   The sequence is that of the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
XX
SQ   Sequence   1721 AA;


  Query Match             8.0%;  Score 175;  DB 19;  Length 1721;
  Best Local Similarity   23.2%;  Pred. No. 5.9e-05;
  Matches   96;  Conservative   41;  Mismatches  180;  Indels   96;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
Db     326 ttttttttttkkptttttttttttttttttttttttt-------ttttttttttttt 378

Qy     103 KSSIKAFGTVTL----PIAFNVGGTGSSTDLEDS--------------KCFTAGTNTVTF 144
Db     379 kpttttttttttttkkptttttattttttsetesvikpdewcwlekngeceakgatyvgv 438

Qy     145 NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 204
Db     439 igkdgriengmaftmipnddthvrfrfkvkdvgntisv-----rcrkg--agklefpdrs 491

Qy     205 GDVAI-------DCS---------NIHIGI--TKGLNDWNYPVS-SESFS--YTKTCTSN 243
Db     492 ldftippvaghnscsiivgvsgdgkihvspygskdvslisapiqpselfnevycdtcta- 550

Qy     244 GIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSD 303
Db     551 ----kygaihsgyqtsadfvtttt--------------------akpttttgapgqpt 585

Qy     304 AGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIITTSYVGVTTSYL 363
Db     586 ttttgspskpttttttkatttttttl--npilttt------tqkpttttttkvpg------ 631

Qy     364 TKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
Db     632 --kppiatttttlkpivttttttkatttttttvptttt-ttkrdemtttttplp 681


RESULT   5
AAB11727
ID   AAB11727 standard; Protein; 1721 AA.
XX
AC   AAB11727;
XX
DT   28-OCT-2000  (first entry)
XX
DE   Portion of Cryptosporidium parvum NINC isolate GP900.
XX
KW   GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide.
XX
OS   Cryptosporidium parvum.
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-0928361.
XX
PR   13-SEP-1996;   96US-0026062.
PR   01-JUN-1993;   93US-0071880.
PR   29-MAY-1992;   92US-0891301.
PR   03-APR-1995;   95US-0415751.
PR   14-AUG-1996;   96US-0700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   N-PSDB; AAA61848, AAA61849.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and
PT   diagnosing or treating Cryptosporidium infections by competitive
PT   inhibition of the function of GP900 -
XX
PS   Claim 2; Column 61-70; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents a portion of the GP900 protein of the NINC isolate of
CC   Cryptosporidium parvum.
XX
SQ   Sequence   1721 AA;


  Query Match             8.0%;  Score 175;  DB 21;  Length 1721;
  Best Local Similarity   23.2%;  Pred. No. 5.9e-05;
  Matches   96;  Conservative   41;  Mismatches  180;  Indels   96;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
Db     326 ttttttttttkkptttttttttttttttttttttttt-------ttttttttttttt 378

Qy     103 KSSIKAFGTVTL----PIAFNVGGTGSSTDLEDS--------------KCFTAGTNTVTF 144
Db     379 kpttttttttttttkkptttttattttttsetesvikpdewcwlekngeceakgatyvgv 438

Qy     145 NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 204
Db     439 igkdgriengmaftmipnddthvrfrfkvkdvgntisv-----rcrkg--agklefpdrs 491

Qy     205 GDVAI-------DCS---------NIHIGI--TKGLNDWNYPVS-SESFS--YTKTCTSN 243
Db     492 ldftippvaghnscsiivgvsgdgkihvspygskdvslisapiqpselfnevycdtcta- 550

Qy     244 GIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSD 303
Db     551 ----kygaihsgyqtsadfvtttt--------------------akpttttgapgqpt 585

Qy     304 AGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIITTSYVGVTTSYL 363
Db     586 ttttgspskpttttttkatttttttl--npilttt------tqkpttttttkvpg------ 631

Qy     364 TKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
```

```
Db    632 --kppiatttttlkpivtttttkatttttttvptttt-ttkrdemtttttplp 681


RESULT   6
ABB30137
ID   ABB30137 standard; Peptide; 688 AA.
XX
AC   ABB30137;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Peptide #2788 encoded by breast cell single exon nucleic acid probe.
XX
KW   Human; microarray; single exon probe; gene expression; breast;
KW   disease; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00662.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes,
PT   useful for measuring gene expression in sample derived from human
PT   breast, comprises number of single exon nucleic acid probes -
XX
PS   Claim 27; SEQ ID NO 13105; 327pp + sequence listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting
CC   the probes with a collection of detectably labelled nucleic acids
CC   derived from mRNA of human breast, and then measuring the label
CC   bound to each probe of the microarray. The probes are useful for
CC   verifying the expression of regions of genomic DNA predicted to
CC   encode proteins. They are useful for gene discovery, and for
CC   determining predisposition and/or prognosing breast disease. Gene
CC   expression analysis is useful for assessing the toxicity of chemical
CC   agents on cells. The microarray of this invention presents a far greater
CC   diversity of probes for measuring gene expression, with far less bias
CC   than expressed sequence tag microarrays. The method is suitable for
CC   rapid production of functional information from genomic sequence. The
CC   present sequence is a peptide encoded by a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   688 AA;


  Query Match             7.1%;  Score 156;  DB 22;  Length 688;
  Best Local Similarity   22.7%;  Pred. No. 0.00055;
  Matches   93;  Conservative   64;  Mismatches  176;  Indels   76;  Gaps   16;

Qy     43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG----------VKYATCQFYSGEEFTTF 91
Db      6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy     92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
Db     56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy    152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
Db    113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy    211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
Db    167 --------tagse--tipastagsettttstegsetttastegs-----etttastess 211

Qy    271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
Db    212 etttattigsetttastegsettttstegsetttastegseittvsttgsetttastegs 271

Qy    325 ----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
Db    272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy    368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
Db    328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT   7
ABB35307
ID   ABB35307 standard; Peptide; 688 AA.
XX
AC   ABB35307;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #2813 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID NO 27942; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
```

```
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   688 AA;


  Query Match             7.1%;  Score 156;  DB 22;  Length 688;
  Best Local Similarity   22.7%;  Pred. No. 0.00055;
  Matches   93;  Conservative   64;  Mismatches  176;  Indels   76;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG----------VKYATCQFYSGEEFTTF 91
           ||    :| |::       | |:|:|     |           | :|    :| | ||
Db       6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy      92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
           ||    :  |   ||    | |      :   | |:   | :   | |: | | :    :
Db      56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy     152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
           ::  :   ::|:  :     ::|  : :|:||:|     ||  |:     :|    |:
Db     113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy     211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
               | |      | |:     | | :: | :    :         :   ::|: :
Db     167 --------tagse--tipastagsettttstegsetttastegs-----etttastess 211

Qy     271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
           : | | |    |   |   |:  |   | : : | : |   |   : | :::|||
Db     212 etttattigsetttastegsettttstegsetttastegsettasttgsetttastegs 271

Qy     325 ----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                    :|||:    |   ::|| :  :   |||:||     ||     |  |  :
Db     272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy     368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
             | || |  ::    || : |   | |::|| | |    |:  :|  |
Db     328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT   8
ABB20749
ID   ABB20749 standard; Protein; 688 AA.
XX
AC   ABB20749;
XX
DT   23-JAN-2002  (first entry)

DE   Protein #2748 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression; heart; microarray; vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 15; SEQ ID No 22519; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays.
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   688 AA;


  Query Match             7.1%;  Score 156;  DB 22;  Length 688;
  Best Local Similarity   22.7%;  Pred. No. 0.00055;
  Matches   93;  Conservative   64;  Mismatches  176;  Indels   76;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG----------VKYATCQFYSGEEFTTF 91
           ||    :| |::       | |:|:|     |           | :|    :| | ||
Db       6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy      92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
           ||    :  |   ||    | |      :   | |:   | :   | |: | | :    :
Db      56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy     152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
           ::  :   ::|:  :     ::|  : :|:||:|     ||  |:     :|    |:
Db     113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy     211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
               | |      | |:     | | :: | :    :         :   ::|: :
Db     167 --------tagse--tipastagsettttstegsetttastegs-----etttastess 211

Qy     271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
           : | | |    |   |   |:  |   | : : | : |   |   : | :::|||
Db     212 etttattigsetttastegsettttstegsetttastegseittvsttgsetttastegs 271

Qy     325 ----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                    :|||:    |   ::|| :  :   |||:||     ||     |  |  :
Db     272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy     368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
             | || |  ::    || : |   | |::|| | |    |:  :|  |
Db     328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT   9
AAM56138
ID   AAM56138 standard; Protein; 688 AA.
XX
AC   AAM56138;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 28243.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
```

PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 28243; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 688 AA;

Query Match 7.1%; Score 156; DB 22; Length 688;
Best Local Similarity 22.7%; Pred. No. 0.00055;
Matches 93; Conservative 64; Mismatches 176; Indels 76; Gaps 16;

```
Qy     43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG-----------VKYATCQFYSGEEFTTF 91
          ||    :| |::       | |:|:|     |           | :|    :| | ||
Db      6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy     92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
          ||    :  |   ||    | |      :   | |:   | :   | |: | | :    :
Db     56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy    152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
          ::  :   ::|:  :     ::|  : :|:||:|     ||  |:     :|    |:
Db    113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy    211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
              | |      | |:     | | :: | :    :          :   ::|: :
Db    167 --------tagse--tipastagsettttstegsetttastegs-----etttastess 211

Qy    271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
          : | | |    |   |   |: |    | : : | : |  |   : | :::|||
Db    212 etttattigsetttastegsettttstegsetttastegseittvsttgsetttastegs 271

Qy    325 -----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                     :||:    |   ::|| :  :   |||:||     ||     |  |  :
Db    272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy    368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
            | || |  ::    || : |   | |::|| | |    |:  :|  |
Db    328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376
```

RESULT 10
AAM68511
ID AAM68511 standard; Protein; 688 AA.
XX
AC AAM68511;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28817.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 28817; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 688 AA;

Query Match 7.1%; Score 156; DB 22; Length 688;
Best Local Similarity 22.7%; Pred. No. 0.00055;
Matches 93; Conservative 64; Mismatches 176; Indels 76; Gaps 16;

```
Qy     43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG-----------VKYATCQFYSGEEFTTF 91
          ||    :| |::       | |:|:|     |           | :|    :| | ||
Db      6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy     92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
          ||    :  |   ||    | |      :   | |:   | :   | |: | | :    :
Db     56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy    152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
          ::  :   ::|:  :     ::|  : :|:||:|     ||  |:     :|    |:
Db    113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy    211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
              | |      | |:     | | :: | :    :          :   ::|: :
Db    167 --------tagse--tipastagsettttstegsetttastegs-----etttastess 211

Qy    271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
          : | | |    |   |   |: |    | : : | : |  |   : | :::|||
Db    212 etttattigsetttastegsettttstegsetttastegseittvsttgsetttastegs 271

Qy    325 -----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                     :||:    |   ::|| :  :   |||:||     ||     |  |  :
Db    272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327
```

```
Qy    368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
          | || |  ::    || : |   | |::|| | |    |:  :|  |
Db    328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT 11
AAM16315
ID   AAM16315 standard; Protein; 688 AA.
XX
AC   AAM16315;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #2749 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 21141; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   688 AA;


  Query Match             7.1%;  Score 156;  DB 22;  Length 688;
  Best Local Similarity   22.7%;  Pred. No. 0.00055;
  Matches   93;  Conservative   64;  Mismatches  176;  Indels   76;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG----------VKYATCQFYSGEEFTTF 91
           ||    :| |::       | |:|:|     |           | :|    :| | ||
Db       6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy      92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
           ||    :  |   ||    | |     :   | |:   | :   | |: | | :    :
Db      56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy     152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
           ::  :   ::|:  :     ::|  : :|:||:|     ||  |:     :|    |:
Db     113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy     211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
                 | |      | |:     | | :: | :    :         :   ::|: :
Db     167 --------tagse--tipastagsettttstegsetttastegs-----etttastess 211

Qy     271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
           : | | |    |   |   |:  |    | : : | : |  |  : | :::|||
Db     212 etttattigsetttastegsettttstegsetttastegseittvsttgsetttastegs 271

Qy     325 ----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                    :||:    |   ::|| :  :   |||:||     ||     |  |  :
Db     272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy     368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
          | || |  ::    || : |   | |::|| | |    |:  :|  |
Db     328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT 12
AAM28810
ID   AAM28810 standard; Protein; 688 AA.
XX
AC   AAM28810;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #2847 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 29079; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   688 AA;


  Query Match             7.1%;  Score 156;  DB 22;  Length 688;
  Best Local Similarity   22.7%;  Pred. No. 0.00055;
```

```
  Matches   93;  Conservative   64;  Mismatches  176;  Indels   76;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG----------VKYATCQFYSGEEFTTF 91
           ||    :| |::       | |:|:|     |            | :|    :| | ||
Db       6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy      92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
           ||    :  |   ||    | |     :   | |:   | :   | |: | | :    :
Db      56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy     152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
           ::  :   ::|:  :      ::|  : :|:||:|    ||  |:     :|     |:
Db     113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy     211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
               | |      | |:     | | :: | :    :          :   ::|: :
Db     167 --------tagse--tipastagsettttttstegsetttastegs-----etttastess 211

Qy     271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
           : | | |    |   |   |:  |    | : : | : |  |   : | :::|||
Db     212 etttattigsetttastegsetttstegsetttastegseittvsttgsetttastegs 271

Qy     325 -----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                      :||:    |   ::|| :  :   |||:||     ||     |  | :
Db     272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy     368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
             | || |  ::    || : |   | |::|| | |    |:  :|  |
Db     328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT  13
AAM04053
ID   AAM04053 standard; Protein; 688 AA.
XX
AC   AAM04053;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Peptide #2735 encoded by probe for measuring breast gene expression.
XX
KW   Probe; human; breast disease; breast cancer; development disorder;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 27; SEQ ID No 12793; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes
CC   (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for measuring human gene expression in
CC   a human breast sample, where the probe hybridises at high stringency to a
CC   nucleic acid expressed in the human breast. The probes are useful for
CC   predicting, diagnosing, grading, staging, monitoring and prognosing
CC   diseases of the human breast, particularly those diseases with polygenic
CC   aetiology. The diseases include: breast cancer, disorders of development,
CC   inflammatory diseases of the breast, fibrocystic changes, proliferative
CC   breast disease and non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   688 AA;


  Query Match             7.1%;  Score 156;  DB 22;  Length 688;
  Best Local Similarity   22.7%;  Pred. No. 0.00055;
  Matches   93;  Conservative   64;  Mismatches  176;  Indels   76;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADG----------VKYATCQFYSGEEFTTF 91
           ||    :| |::       | |:|:|     |            | :|    :| | ||
Db       6 tstegsetttvsa------tgsettvstegsgtttvsitgsettkvst----tgsetttt 55

Qy      92 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 151
           ||    :  |   ||    | |     :   | |:   | :   | |: | | :    :
Db      56 stegseitta---sitgsetttastegsetttastegsettsasttgsetttasttsset 112

Qy     152 SI-DVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAID 210
           ::  :   ::|:  :      ::|  : :|:||:|    ||  |:     :|     |:
Db     113 tmasimgsettmastigsettkvstasskmttvft----ensettiasttasetttvs-- 166

Qy     211 CSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN 270
               | |      | |:     | | :: | :    :          :   ::|: :
Db     167 --------tagse--tipastagsettttttstegsetttastegs-----etttastess 211

Qy     271 QYTLAYT-NDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA----- 324
           : | | |    |   |   |:  |    | : : | : |  |   : | :::|||
Db     212 etttattigsetttastegsetttstegsetttastegseittvsttgsetttastegs 271

Qy     325 -----------VTTLPFNPSVDKTKTIEI-LQPIPTTTITTSYVGVTT-----SYLTKTA 367
                      :||:    |   ::|| :  :   |||:||     ||     |  | :
Db     272 etttastegselttv----sttgsetitvsaegsetttvttmgsetttastagsetttvs 327

Qy     368 PIG-ETATVIVDVPYHTTTTVTSEWTGTITTTTTRT--NPTDSIDTVVV 413
             | || |  ::    || : |   | |::|| | |    |:  :|  |
Db     328 tagsetttasiegsetttvsstgsetttvsttgtettitstegsetttv 376


RESULT  14
AAR47578
ID   AAR47578 standard; Protein; 894 AA.
XX
AC   AAR47578;
XX
DT   19-JUL-1994  (first entry)
XX
DE   Flocculation protein of Saccharomyces cerevisiae.
XX
KW   Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW   Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW   anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW   inulinase; alpha-amylase; Saccharomyces cerevisiae;
KW   floculation protein; enzymatic process; fermentation;
KW   biodegradation; catalysis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO9401567-A.
XX
PD   20-JAN-1994.
XX
PF   07-JUL-1993;   93WO-EP01763.
XX
```

PR 08-JUL-1992; 92EP-0202080.
PR 14-DEC-1992; 92EP-0203899.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
PI Klis FM, Schreuder MP, Toschka H, Verrips CT;
XX
DR WPI; 1994-035071/04.
DR N-PSDB; AAQ54029.
XX
PT Immobilisation of enzymes to microbial cell wall - by prodn. of
PT fusion protein of enzyme linked to anchoring protein
XX
PS Example 10; Page 59-64; 99pp; English.
XX
CC The flocculation protein is used in a method to immobilise enzymes
CC to a microbial cell wall. The coding sequence (FLO 1) is used in
CC the production of a recombinant polynucleotide which comprises a
CC structural gene encoding a protein with catalytic activity and at
CC least part of a gene encoding at least the C-terminus of a protein
CC capable of anchoring in a eukaryotic or prokaryotic cell wall. The
CC anchoring fragment or protein is selected from alpha agglutinin,
CC AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a
CC proteinase of lactic acid bacteria. The recombinant polynucleotide
CC preferably also comprises a sequence encoding a signal peptide to
CC ensure secretion of the expressed product. The signal peptide is
CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC alpha-amylase of Bacillus or proteinases of lactic acid bacteria.
CC The host microorganism can be used for performing enzymatic
CC processes on an industrial scale.
XX
SQ Sequence 894 AA;

Query Match 7.1%; Score 155; DB 15; Length 894;
Best Local Similarity 23.3%; Pred. No. 0.00096;
Matches 109; Conservative 49; Mismatches 181; Indels 128; Gaps 22;

```
Qy      12 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCV------------ 58
           :| |:|||| |:   || :       |         :   |:|||
Db      50 dsstysnaaymay---gyaskt-----klgsvggqtdisidynipcvsssgtfpcpqeds 101

Qy      59 ------------------------FKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFST 93
                                   | : |: |:| |   |       ::   :  :::
Db     102 ygnwgckgmgacsnsqgiaywstdlfgfyttptnvtlemtg-------yflppqtgsytf 154

Qy      94 LTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI-- 151
             ||:|   |:| : |          | |  :   :     |:   |:   || |
Db     155 kfatvdd---sailsvg----------gatafnccaqqqppitstnfti---dgikpwgg 198

Qy     152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVT--TLFVAPQCENGYT----------SGTMG 199
           |:    | :    : | |  :|: | | |:  || ::    :| |          |
Db     199 slppniegtvymyagyyypmkvvys-navswgtlpisvtlpdgttvsddfegyvysfddd 257

Qy     200 FSSSNGDVAIDCSNIHIGITKGLND-WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRP 258
            | ||  |  | ||  :  |    : |    :| |   |    :||:    : |     |
Db     258 lsqsnctvp-dpsnyavstttttepwtgtftststemttvtgtngvptd-etvivirtp 315

Qy     259 FIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTV 318
             :  || |                          |  |||   |   :    |  |
Db     316 tseglistt-------------------------tepwtgtfts-tstevttitgtngqp 349

Qy     319 TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI-V 377
           || |  |   | :  :  | |    :|  | | |::   :||   |   |  ||  ||
Db     350 tdetvivirtptseglisttt----epw-tgtftststemttvtgtngqptdetvivirt 404

Qy     378 DVPYHTTTTVTSEWTGTITTTTTR------TN--PTDSIDTVVVQVP 416
                 || |  |||| |:|:|       ||  |||    :||: |
Db     405 ptseglvtttepwtgtftststemstvtgtnglptde-tvivvktp 450
```

RESULT 15
AAR58754
ID AAR58754 standard; Protein; 894 AA.
XX
AC AAR58754;
XX
DT 27-MAR-1995 (first entry)
XX
DE S. cerevisiae FLO1.
XX
KW Binding protein; immobilization; chimeric protein;
KW anchoring protein; Saccharomyces cerevisiae; flocculation;
KW FLO1 gene; scFv; single chain antibody; monoclonal antibody;
KW MAb; human chorionic gonadotropin; HCG.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9418330-A.
XX
PD 18-AUG-1994.
XX
PF 10-FEB-1994; 94WO-EP00427.
XX
PR 10-FEB-1993; 93EP-0200350.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
PI De Geus P, Frenken LGJ, Klis FM, Toschka HY, Verrips CT;
XX
DR WPI; 1994-279751/34.
DR N-PSDB; AAQ67360.
XX
PT Immobilised binding proteins for specific cpds - obtd. by
PT expressing chimeric proteins comprising the binding protein and a
PT cell wall-anchoring protein in host cells
XX
PS Disclosure; Page 36-39; 78pp; English.
XX
CC Valuable compounds are isolated from complex mixtures by use of
CC immobilized ligands composed of an anchoring protein and a binding
CC protein. A suitable anchoring protein is yeast FLO1 (associated
CC with flocculation), and a gene encoding a chimeric scFv-FLO1 protein
CC that will anchor in the cell wall of a lower eukaryote, and which
CC binds HCG with high specificity, was produced.
XX
SQ Sequence 894 AA;

Query Match 7.1%; Score 155; DB 15; Length 894;
Best Local Similarity 23.3%; Pred. No. 0.00096;
Matches 109; Conservative 49; Mismatches 181; Indels 128; Gaps 22;

```
Qy      12 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCV------------ 58
           :| |:|||| |:   || :       |  |     :   |:|||
Db      50 dsstysnaaymay---gyaskt-----klgsvggqtdisidynipcvsssgtfpcpqeds 101

Qy      59 ------------------------FKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFST 93
                                   | : |: |:| |   |       ::   :  :::
Db     102 ygnwgckgmgacsnsqgiaywstdlfgfyttptnvtlemtg-------yflppqtgsytf 154

Qy      94 LTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI-- 151
             ||:|   |:| : |          | |  :   :     |:   |:   || |
Db     155 kfatvdd---sailsvg----------gatafnccaqqqppitstnfti---dgikpwgg 198

Qy     152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVT--TLFVAPQCENGYT----------SGTMG 199
           |:    | :    : | |  :|: | | |:  || ::    :| |          |
Db     199 slppniegtvymyagyyypmkvvys-navswgtlpisvtlpdgttvsddfegyvysfddd 257

Qy     200 FSSSNGDVAIDCSNIHIGITKGLND-WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRP 258
            | ||  |  | ||  :  |    : |    :| |   |    :||:    : |     |
Db     258 lsqsnctvp-dpsnyavstttttepwtgtftststemttvtgtngvptd-etvivirtp 315
```

```
Qy     259 FIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTV 318
           :  || |                              |  |||   |   :    |  |
Db     316 tseglistt------------------------tepwtgtfts-tstevttitgtngqp 349

Qy     319 TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI-V 377
           || |  |   | :  :  | |    :|  | | |::   :||   |   |  ||  ||
Db     350 tdetvivirtptseglisttt----epw-tgtftststemttvtgtngqptdetvivirt 404

Qy     378 DVPYHTTTTVTSEWTGTITTTTTR------TN--PTDSIDTVVVQVP 416
               || |  |||| |:|:|       ||  |||    :||: |
Db     405 ptseglvtttepwtgtftststemstvtgtnglptde-tvivvktp 450
```

Search completed: October 3, 2002, 16:29:43
Job time: 246 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:27:32 ; Search time 14.64 Seconds
(without alignments)
694.061 Million cell updates/sec

Title: US-09-715-876-8_COPY_17_432
Perfect score: 2190
Sequence: 1 AKTITGVFDSFNSLTWSNAA.....TTTTRTNPTDSIDTVVVQVP 416

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	308.5	14.1	650	3	US-08-362-525-2	Sequence 2, Appli
2	177	8.1	1537	1	US-08-325-267A-2	Sequence 2, Appli
3	175.5	8.0	1837	3	US-08-928-361B-5	Sequence 5, Appli
4	175	8.0	1721	3	US-08-700-651-5	Sequence 5, Appli
5	175	8.0	1721	3	US-08-928-361B-6	Sequence 6, Appli
6	155	7.1	894	3	US-08-362-525-22	Sequence 22, Appl
7	155	7.1	894	3	US-08-971-692-15	Sequence 15, Appl
8	150.5	6.9	175	3	US-08-700-651-12	Sequence 12, Appl
9	150.5	6.9	175	3	US-08-928-361B-17	Sequence 17, Appl
10	148.5	6.8	216	3	US-08-928-361B-8	Sequence 8, Appli
11	148.5	6.8	862	1	US-08-325-267A-4	Sequence 4, Appli
12	146.5	6.7	249	3	US-08-700-651-15	Sequence 15, Appl
13	146.5	6.7	249	3	US-08-928-361B-20	Sequence 20, Appl
14	145	6.6	130	3	US-08-700-651-9	Sequence 9, Appli
15	145	6.6	130	3	US-08-928-361B-14	Sequence 14, Appl
16	143.5	6.6	216	3	US-08-928-361B-27	Sequence 27, Appl
17	140	6.4	130	3	US-08-700-651-8	Sequence 8, Appli
18	140	6.4	130	3	US-08-928-361B-13	Sequence 13, Appl
19	140	6.4	138	3	US-08-700-651-10	Sequence 10, Appl
20	140	6.4	138	3	US-08-928-361B-15	Sequence 15, Appl
21	139.5	6.4	150	3	US-08-928-361B-18	Sequence 18, Appl
22	139	6.3	128	3	US-08-700-651-7	Sequence 7, Appli
23	139	6.3	128	3	US-08-928-361B-12	Sequence 12, Appl
24	138.5	6.3	162	3	US-08-700-651-13	Sequence 13, Appl
25	136	6.2	1477	1	US-08-038-682-4	Sequence 4, Appli
26	136	6.2	1477	1	US-08-302-832-4	Sequence 4, Appli
27	136	6.2	1477	2	US-08-530-198-4	Sequence 4, Appli
28	136	6.2	1477	2	US-08-469-880-4	Sequence 4, Appli
29	136	6.2	1477	2	US-08-728-470-4	Sequence 4, Appli
30	136	6.2	1477	2	US-08-617-697-4	Sequence 4, Appli
31	136	6.2	1477	4	US-08-719-641-4	Sequence 4, Appli
32	134.5	6.1	124	3	US-08-700-651-11	Sequence 11, Appl
33	134.5	6.1	124	3	US-08-928-361B-16	Sequence 16, Appl
34	134	6.1	1222	2	US-08-682-517-15	Sequence 15, Appl
35	134	6.1	1252	2	US-08-682-517-9	Sequence 9, Appli
36	132	6.0	629	4	US-09-241-581B-6	Sequence 6, Appli
37	132	6.0	629	5	PCT-US95-07721-6	Sequence 6, Appli
38	131.5	6.0	562	6	5258502-2	Patent No. 5258502
39	128	5.8	907	3	US-08-783-774-2	Sequence 2, Appli
40	128	5.8	907	5	PCT-US95-04611A-19	Sequence 19, Appl
41	127.5	5.8	1426	4	US-09-136-574A-43	Sequence 43, Appl
42	126.5	5.8	451	1	US-08-287-001A-2	Sequence 2, Appli
43	126.5	5.8	451	5	PCT-US95-09941-2	Sequence 2, Appli
44	120	5.5	806	1	US-08-270-076A-11	Sequence 11, Appl
45	119.5	5.5	2314	4	US-09-268-347-49	Sequence 49, Appl

ALIGNMENTS

```
RESULT 1
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910
;  GENERAL INFORMATION:
;    APPLICANT:  KLIS, FRANCISCUS M.
;    APPLICANT:  SCHREUDER, MAARTEN P.
;    APPLICANT:  TOSCHKA, HOLSER Y.
;    APPLICANT:  VERRIPS, CORNELIS T.
;    TITLE OF INVENTION:  PROCESS FOR IMMOBILIZING ENZYMES TO THE
;    TITLE OF INVENTION:  CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,525
;      FILING DATE:  04-JAN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92202080.5
;      FILING DATE:  08-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92203899.7
;      FILING DATE:  14-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP93/01763
;      FILING DATE:  07-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  213289/T7020(V)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 amino acids
;      TYPE:  amino acid
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-362-525-2


  Query Match             14.1%;  Score 308.5;  DB 3;  Length 650;
  Best Local Similarity   25.8%;  Pred. No. 1.8e-18;
  Matches  116;  Conservative   79;  Mismatches  190;  Indels   65;  Gaps   20;

Qy      10 SFNSLTWSN------AANYAFKGPGYPTWNAVLGWSL-DGTSANPGDTFTLNMPCVFK-- 60
           : | :|:||        ||   | |    | |   :|: | :|   || |||:|| |::
Db      21 NINDITFSNLEITPLTAN---KQPD-QGWTATFDFSIADASSIREGDEFTLSMPHVYRIK 76

Qy      61 -YTTSQTSVDLTADGVKYATCQFYSGE-----EFTTFSTLTCTVNDALKSSIKAFGTVTL 114
             :|||:    ||| :   | : | :     | |||   |||  : | |     |::|
Db      77 LLNSSQTATISLADGTEAFKC-YVSQQAAYLYENTTF---TCTAQNDLSSYNTIDGSITF 132

Qy     115 PIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY---LYAS 171
           : |: ||:    :||::| | :|   |   :   |:   | |     ||:|:   :: |
Db     133 SLNFSDGGSSYEYELENAKFFKSGPMLVKLGNQMSDV---VNF-----DPAAFTENVFHS 184

Qy     172 RVMPSLNKVTTLFVAPQCENGYTSG---TMGFSSSNGDVAIDCSNIHIGITKGLNDWNYP 228
           :   :   | |||   |    : : ||| :| :|||:: :   :    ||| :|
Db     185 GRSTGYGSFESYHLGMYCPNGYFLGGTEKIDYDSSNNNVDLDCSSVQVYSSNDFNDWWFP 244

Qy     229 VSSESFSYTKTCTSNGIQIKY-QNVPAGYRPFIDAYIS-ATDVNQYTLAYTNDYTC---- 282
           |    :  ||  : : |   : : |   :::|  |   :||    |   |||
Db     245 QSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNALQSLPANVNTIDHALEFQYTCLDTI 304

Qy     283 AGSRLQSKPFTLR----WTGYKNSDAGSNGIVIVATTRTVTD------STTAVTTLPFNP 332
           | :   ::  | |    : |     | :    |  ||  :|       || :::|:
Db     305 ANTTYATQFSTTREFIVYQGRNLGTASAKSSFISTTTTDLTSINTSAYSTGSISTV---- 360

Qy     333 SVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTT------- 385
           | |::  : ||:   |   |:  : :|:     : : |    |||:
Db     361 ETGNRTTSEVISHVVTTSTKLSPTATTSLTIAQTSIYSTDSNITVGTDIHTTSEVISDVE 420

Qy     386 TVTSEWTGTITTT-TTRTNPTDSIDTVVVQ 414
           |:: |   |:    |: |  | :::| : |
Db     421 TISRETASTVVAAPTSTTGWTGAMNTYIPQ 450


RESULT   2
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1537 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-267A-2


  Query Match             8.1%;  Score 177;  DB 1;  Length 1537;
  Best Local Similarity   23.0%;  Pred. No. 1.5e-06;
  Matches  112;  Conservative   47;  Mismatches  181;  Indels  146;  Gaps   25;

Qy      21 NYAFKGPG-YPTWNAVLGWSLD--GTSANPGDTFTLNM----------PCVFKYTTSQ-- 65
           |: || |       :  || |  |    | :  || |             ||: |
Db     104 NWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDS 162

Qy      66 ---------------------TSVDLTADGVK----------YATCQFYSGEEFTTFSTL 94
                                 || : | ||:|           |   |:|  :
Db     163 AILSVGGATAFNCCAQQQPPITSTNFTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---- 218

Qy      95 TCTVNDALKSSIKAFGT----VTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKD 150
                 : |:  ::||    ||||       ||  | | |             :|:|
Db     219 -----KVVYSNAVSWGTLPISVTLP-----DGTTVSDDFEG--------YVYSFDD---- 256

Qy     151 ISIDVEFEKSTV-DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI 209
             |:    || ||| |   |    :    |  | :   |    :|| |  :    : |
Db     257 ---DLSQSNCTVPDPSNYA-VSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVI 312

Qy     210 DCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDV 269
                 |      ||   :| | |    |    :||::      :       |    :||
Db     313 RTPTTASTIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETII------VIRTPTTATTA 366

Qy     270 NQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVT 326
             |  : : :|   : |                 :  |:||:     :   || | :|||:|
Db     367 ITTTEPWNSTFTSTSTEL------------TTVTGTNGLPTDETIIVIRTPTTATTAMT 413

Qy     327 TL-PFNPSVDKTKT----------------IEILQPIPTTTITT-------SYVGVTTS 361
           |  |:| :   | |                 | |  |   ||  |       ::    :|
Db     414 TTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTE 473

Qy     362 YLTKTA----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTTTRTN--PTDS 407
             | |     |  ||  :::  |   ||  | |  | | |       || | ||  |||
Db     474 MTTVTGTNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTD- 531

Qy     408 IDTVVV 413
             :|::|
Db     532 -ETIIV 536


RESULT   3
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
```

```
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1837 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-5


  Query Match             8.0%;  Score 175.5;  DB 3;  Length 1837;
  Best Local Similarity   22.9%;  Pred. No. 2.7e-06;
  Matches   93;  Conservative   45;  Mismatches  176;  Indels   93;  Gaps   13;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
           |:  |  | |          ||: |:   |       |    : :: || :| | |
Db     452 TTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTKKPTTTTTTTTTTTKK 511

Qy     103 KSSIKAFGTVTLPIAFNVGGTGSSTDL----------EDSKCFTAGTNTVTFNDGDKDI 151
             ::     | |          | | |:          :: :|   |   |     |  |
Db     512 PTTTTTATTTT---------TTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRI 562

Qy     152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI-- 209
             : |     | :   :  :|    | ::      :|  |  :| : |   : |  |
Db     563 ENGMAFTMIPNDDTHVRFRFKVKDVGNTISV-----RCGKG--AGKLEFPDRSLDFTIPP 615

Qy     210 -----DCSNIHIGITKGLNDWNYPVSSESFS--------------YTKTCTSNGIQIKY 249
                 || | :|:: |      |  |:  |               |  |||:     ||
Db     616 VAGHNSCS-IIVGVSGGGKIHVSPYGSKDVSLISAPIQPCELFNEVYCDTCTA-----KY 669

Qy     250 QNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI 309
             : :||:   |   : |                    :|| |        :  |: |
Db     670 GAIHSGYQTSADFVTTTT--------------------AKPTT--------TTTGAPGQ 700

Qy     310 VIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIITTSYVGVTTSYLTKTAPI 369
             || : :  ||  ||         | |  || || ||| |      ||: :    ||
Db     701 PTTTTTGSPSKPTTTTTT-------KATTTTTILNPIITTT-TQKPTTTTTTKVPGKPPI 752

Qy     370 GETATVIVDVPYHTTTTVTSEWTGTITTTTTTRTNPTDSIDTVVVQVP 416
           | | :  :   |||  |:  | |: |||| |   | : |    :|
Db     753 ATTTTTLKPIVTTTTTKATTTTTTTVPTTTT-TTKRDEMTTTTTPLP 798


RESULT   4
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1721
;   TYPE: PRT
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-5


  Query Match             8.0%;  Score 175;  DB 3;  Length 1721;
  Best Local Similarity   23.2%;  Pred. No. 2.7e-06;
  Matches   96;  Conservative   41;  Mismatches  180;  Indels   96;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
           |:      | |   |     ||: |:    |       |          || :| | |
Db     326 TTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTT-------TTTTTTTTTTTTT 378

Qy     103 KSSIKAFGTVTL----PIAFNVGGTGSSTDLEDS-------------KCFTAGTNTVTF 144
           | :     | |     |       | ::|   :|              :|   |   |
Db     379 KPTTTTTTTTTTTTKKPTTTTTATTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGV 438

Qy     145 NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 204
             |  |   : |     | :   :  :|    | ::      :|  |  :| : |   :
Db     439 IGKDGRIENGMAFTMIPNDDTHVRFRFKVKDVGNTISV-----RCRKG--AGKLEFPDRS 491

Qy     205 GDVAI-------DCS---------NIHIGI--TKGLNDWNYPVS-SESFS--YTKTCTSN 243
             | |           ||         ||:    :| ::  : |:  || |:  |  |||:
Db     492 LDFTIPPVAGHNSCSIIVGVSGDGKIHVSPYGSKDVSLISAPIQPSELFNEVYCDTCTA- 550

Qy     244 GIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSD 303
             ||  : :||:   |   : |                    :|| |
Db     551 ----KYGAIHSGYQTSADFVTTTT--------------------AKPTTTTTGAPGQPT 585

Qy     304 AGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIITTSYVGVTTSYL 363
             : |     || | | :||  |||  || :  |         |||| ||   |
Db     586 TTTTGSPSKPTTTTTTKATTTTTTL--NPIITTT------TQKPTTTTTTKVPG------ 631

Qy     364 TKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTTRTNPTDSIDTVVVQVP 416
             ||  | | :  :   |||  |:  | |: |||| |   | : |    :|
Db     632 --KPPIATTTTTLKPIVTTTTTKATTTTTTTVPTTTT-TTKRDEMTTTTTPLP 681


RESULT   5
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
```

```
;       ZIP:  94306-1840
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/928,361B
;       FILING DATE:  12-SEP-1997
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/026,062
;       FILING DATE:  13-SEP-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Verny, Hana
;       REGISTRATION NUMBER:  30,518
;       REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-324-1677
;       TELEFAX:  650-324-1678
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1721 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-928-361B-6


  Query Match             8.0%;  Score 175;  DB 3;  Length 1721;
  Best Local Similarity   23.2%;  Pred. No. 2.7e-06;
  Matches   96;  Conservative   41;  Mismatches  180;  Indels   96;  Gaps   16;

Qy      43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
           |:     | |   |     ||: |:   |       |         || :| | |
Db     326 TTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTT-------TTTTTTTTTTTTT 378

Qy     103 KSSIKAFGTVTL----PIAFNVGGTGSSTDLEDS--------------KCFTAGTNTVTF 144
           | :     | |     |       | ::|   :|              :|   |   |
Db     379 KPTTTTTTTTTTTTKKPTTTTATTTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGV 438

Qy     145 NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 204
             |  |   : |     | :   :  :|    | ::      :|  |  :| : |   :
Db     439 IGKDGRIENGMAFTMIPNDDTHVRFRFKVKDVGNTISV-----RCRKG--AGKLEFPDRS 491

Qy     205 GDVAI-------DCS---------NIHIGI--TKGLNDWNYPVS-SESFS--YTKTCTSN 243
             |  |         ||          ||:    :| ::  : |:  || |:  |  |||:
Db     492 LDFTIPPVAGHNSCSIIVGVSGDGKIHVSPYGSKDVSLISAPIQPSELFNEVYCDTCTA- 550

Qy     244 GIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSD 303
               ||  : :||:   |   : |                  :|| |
Db     551 ----KYGAIHSGYQTSADFVTTTT--------------------AKPTTTTTGAPGQPT 585

Qy     304 AGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYL 363
             : |     || | | :||  |||  || :  |          |||| ||   |
Db     586 TTTTGSPSKPTTTTTTKATTTTTTL--NPIITTT------TQKPTTTTTTKVPG------ 631

Qy     364 TKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
             ||  | | :  :   |||  |:  | |: ||||  |   | : |    :|
Db     632 --KPPIATTTTTLKPIVTTTTTKATTTTTTVPTTTT-TTKRDEMTTTTPLP 681


RESULT   6
US-08-362-525-22
; Sequence 22, Application US/08362525
; Patent No. 6027910
;  GENERAL INFORMATION:
;     APPLICANT:  KLIS, FRANCISCUS M.
;     APPLICANT:  SCHREUDER, MAARTEN P.
;     APPLICANT:  TOSCHKA, HOLSER Y.
;     APPLICANT:  VERRIPS, CORNELIS T.
;     TITLE OF INVENTION:  PROCESS FOR IMMOBILIZING ENZYMES TO THE
;     TITLE OF INVENTION:  CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  26
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;       STREET:  1100 New York Avenue, N.W.
;       CITY:  Washington
;       STATE:  D. C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005-3918
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/362,525
;       FILING DATE:  04-JAN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 92202080.5
;       FILING DATE:  08-JUL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 92203899.7
;       FILING DATE:  14-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/EP93/01763
;       FILING DATE:  07-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOKULIS, PAUL N.
;       REGISTRATION NUMBER:  16,773
;       REFERENCE/DOCKET NUMBER:  213289/T7020(V)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;       TELEX:  6714627 CUSH
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  894 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-362-525-22


  Query Match             7.1%;  Score 155;  DB 3;  Length 894;
  Best Local Similarity   23.3%;  Pred. No. 5.5e-05;
  Matches  109;  Conservative   49;  Mismatches  181;  Indels  128;  Gaps   22;

Qy      12 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCV------------ 58
           :| |:|||  |:   || :          |       :   |:|||
Db      50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDS 101

Qy      59 ------------------------FKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFST 93
                                    | : |: |:| |   |       ::   :  :::
Db     102 YGNWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNVTLEMTG-------YFLPPQTGSYTF 154

Qy      94 LTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI-- 151
             ||:|   |:| : |           | | :   :      |:   |:   || |
Db     155 KFATVDD---SAILSVG----------GATAFNCCAQQQPPITSTNFTI---DGIKPWGG 198

Qy     152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVT--TLFVAPQCENGYT----------SGTMG 199
           |:    | :    : | |  :|: | | |:  || ::     :| |          |
Db     199 SLPPNIEGTVYMYAGYYYPMKVVYS-NAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDD 257

Qy     200 FSSSNGDVAIDCSNIHIGITKGLND-WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRP 258
             | ||  |  | ||  :  |    : |    :| |   |    :||:    : |     |
Db     258 LSQSNCTVP-DPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTD-ETVIVIRTP 315

Qy     259 FIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTV 318
             :  || |                          |  |||   |   :    |  |
Db     316 TSEGLISTT------------------------TEPWTGTFTS-TSTEVTTITGTNGQP 349
```

```
Qy    319 TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI-V 377
          || |  |   | :  :  | |    :|  | | |::   :||   |   |  ||  ||
Db    350 TDETVIVIRTPTSEGLISTTT----EPW-TGTFTSTSTEMTTVTGTNGQPTDETVIVIRT 404

Qy    378 DVPYHTTTTVTSEWTGTITTTTTR------TN--PTDSIDTVVVQVP 416
                || |  |||| |:|:|       ||  |||    :||: |
Db    405 PTSEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDE-TVIVVKTP 450


RESULT   7
US-08-971-692-15
; Sequence 15, Application US/08971692
; Patent No. 6114147
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Immobilized proteins with specific binding
;    TITLE OF INVENTION:  capacities and their use in processes and products.
;    NUMBER OF SEQUENCES:  40
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/971,692
;      FILING DATE:
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-971-692-15


  Query Match             7.1%;  Score 155;  DB 3;  Length 894;
  Best Local Similarity   23.3%;  Pred. No. 5.5e-05;
  Matches  109;  Conservative   49;  Mismatches  181;  Indels  128;  Gaps   22;

Qy     12 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCV------------ 58
          :| |:|||| |:   || :       |          :   |:|||
Db     50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDS 101

Qy     59 ------------------------FKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFST 93
                                  | : |: |:| |   |      ::   :  :::
Db    102 YGNWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNVTLEMTG-------YFLPPQTGSYTF 154

Qy     94 LTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI-- 151
             ||:|   |:| : |          | |  :   :     |:   |:   || |
Db    155 KFATVDD---SAILSVG----------GATAFNCCAQQQPPITSTNFTI---DGIKPWGG 198

Qy    152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVT--TLFVAPQCENGYT----------SGTMG 199
          |:    | :    : | |  :|: | | |:  || ::    :| |          |
Db    199 SLPPNIEGTVYMYAGYYYPMKVVYS-NAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDD 257

Qy    200 FSSSNGDVAIDCSNIHIGITKGLND-WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRP 258
           | ||  |  | ||  :  |    : |    :| |   |    :||:    : |     |
Db    258 LSQSNCTVP-DPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTD-ETVIVIRTP 315

Qy    259 FIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTV 318
             :  || |                          |  |||   |   :    |  |
Db    316 TSEGLISTT------------------------TEPWTGTFTS-TSTEVTTITGTNGQP 349

Qy    319 TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI-V 377
          || |  |   | :  :  | |    :|  | | |::   :||   |   |  ||  ||
Db    350 TDETVIVIRTPTSEGLISTTT----EPW-TGTFTSTSTEMTTVTGTNGQPTDETVIVIRT 404

Qy    378 DVPYHTTTTVTSEWTGTITTTTTR------TN--PTDSIDTVVVQVP 416
                || |  |||| |:|:|       ||  |||    :||: |
Db    405 PTSEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDE-TVIVVKTP 450
```

```
RESULT   8
US-08-700-651-12
; Sequence 12, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 175
;   TYPE: PRT
;   ORGANISM: Cryptosporidium parvum
;   FEATURE:
;   OTHER INFORMATION: mutant/variant of SEQ ID NO:5
US-08-700-651-12


  Query Match             6.9%;  Score 150.5;  DB 3;  Length 175;
  Best Local Similarity   32.5%;  Pred. No. 1.2e-05;
  Matches   41;  Conservative   12;  Mismatches   72;  Indels    1;  Gaps    1;

Qy    291 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 350
          |:| :  | |::   :        || | | :||  ||      :  | |       |||
Db      8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 66

Qy    351 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 410
            ||:    ||:  | |     | |         ||||:|:  | | |||||  |  | :  |
Db     67 TTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 126

Qy    411 VVVQVP 416
             : |
Db    127 TTTKKP 132


RESULT   9
US-08-928-361B-17
; Sequence 17, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US 60/026,062
;       FILING DATE:  13-SEP-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Verny, Hana
;       REGISTRATION NUMBER:  30,518
;       REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-324-1677
;       TELEFAX:  650-324-1678
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  175 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-928-361B-17


  Query Match             6.9%;  Score 150.5;  DB 3;  Length 175;
  Best Local Similarity   32.5%;  Pred. No. 1.2e-05;
  Matches   41;  Conservative   12;  Mismatches   72;  Indels    1;  Gaps    1;

Qy     291 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 350
           |:| :  | |::   :      || | | :||  ||     :  | |        |||
Db       8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 66

Qy     351 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 410
             ||:    ||:  | |     | |        |||| |:  | | ||||| |  | :  |
Db      67 TTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 126

Qy     411 VVVQVP 416
               : |
Db     127 TTTKKP 132


RESULT  10
US-08-928-361B-8
; Sequence 8, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  216 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-8


  Query Match             6.8%;  Score 148.5;  DB 3;  Length 216;
  Best Local Similarity   31.5%;  Pred. No. 2.4e-05;
  Matches   46;  Conservative   10;  Mismatches   79;  Indels   11;  Gaps    2;

Qy     265 SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA 324
           : |      |   |   |   :    || |  -|         :       || | | :||
Db      44 TTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTT--------TTTTTTTTTTTTTTTTTT 95

Qy     325 VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT 384
             ||      :   | |    :|  ||| ||:    ||:  | |     | |     |  ||
Db      96 TTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTT---TTTTTTTTTTTTTTTTTTTTKKPTTTT 152

Qy     385 TTVTSEWTGTITTTTTRTNPTDSIDT 410
           || |:  | | ||||| |  | :  |
Db     153 TTTTTTTTTTTTTTTTTTTTTTTTTT 178


RESULT  11
US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
```

```
;     LENGTH: 862 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-267A-4


  Query Match             6.8%;  Score 148.5;  DB 1;  Length 862;
  Best Local Similarity   23.3%;  Pred. No. 0.00019;
  Matches  107;  Conservative   45;  Mismatches  149;  Indels  159;  Gaps   23;

Qy      12 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLT 71
           :| |:|||| |:   || :       |         :   |:|||   ::| |
Db      50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCV---SSSGTFPCPQ 98

Qy      72 ADGVKYATCQ----------FYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIA-- 117
             |    |:              ::| : |  ::| |                 |||  :
Db      99 EDSYGNWGCKGMGACSNSQGIAYWSTDLFGFYTTPT--------------NVTLEMTGY 143

Qy     118 FNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSL 177
           |   ||| |                              |: :||| || |       |:
Db     144 FLPPQTGSYT---------------------------FKFATVDDSAIL-------SV 167

Qy     178 NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDW--NYPVSSESFS 235
           |      | :   ||                :|  |    |:  |  : | : |
Db     168 GGATAFNCCAQQQPPITS---------------TNFTI---DGIKPWGGSLPPNIEGTV 208

Qy     236 YTKTCTSNGIQIKYQN-VPAGYRPFI----DAYISATDVNQYTLAYTND---------- 279
           |       ::: | | |  |  |       |    : |   |  :: :|
Db     209 YMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCTVPDP 268

Qy     280 --YTCAGSRLQSKPFTLRWTGYKNSDA-------GSNGI-----VIVATTRTVTDSTTAV 325
             |  : :   ::|    |||    | :       |:||:     |||  | | | ||
Db     269 SNYAVSTTTTTEP----WTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPT-TASTIIT 323

Qy     326 TTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI-VDVPYHTT 384
           || |:                 | | |::   :||   |   |  ||  ||
Db     324 TTEPW----------------TGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLV 366

Qy     385 TTVTSEWTGTITTTTTR------TN--PTDSIDTVVVQVP 416
           || |  |||| |:|:|       ||  |||    :||: |
Db     367 TTTTEPWTGTFTSTSTEMSTVTGTNGLPTDE-TVIVVKTP 405


RESULT  12
US-08-700-651-15
; Sequence 15, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;   APPLICANT: PETERSEN, CAROLYN
;   APPLICANT: LEECH, JAMES
;   APPLICANT: NELSON, RICHARD, C.
;   APPLICANT: GUT, JIRI
;   TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;   TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;   TITLE OF INVENTION: INFECTIONS
;   FILE REFERENCE: 480.19-4(HV)
;   CURRENT APPLICATION NUMBER: US/08/700,651B
;   CURRENT FILING DATE: 1997-08-14
;   EARLIER APPLICATION NUMBER: 08/415,751
;   EARLIER FILING DATE: 1995-04-03
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;    LENGTH: 249
;    TYPE: PRT
;    ORGANISM: Cryptosporidium parvum
;    FEATURE:
;    OTHER INFORMATION: mutant/variant of SEQ ID NO:5
US-08-700-651-15


  Query Match             6.7%;  Score 146.5;  DB 3;  Length 249;
  Best Local Similarity   33.3%;  Pred. No. 4.4e-05;
  Matches   40;  Conservative   11;  Mismatches   68;  Indels    1;  Gaps    1;

Qy     291 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 350
           |:| :  | |::   :       || | | :||  ||      :   | |        |||
Db       8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 66

Qy     351 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 410
             ||:   ||: | |     | |        ||||  |:  | | |||||  |  | :  |
Db      67 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 126


RESULT  13
US-08-928-361B-20
; Sequence 20, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  249 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-20


  Query Match             6.7%;  Score 146.5;  DB 3;  Length 249;
  Best Local Similarity   33.3%;  Pred. No. 4.4e-05;
  Matches   40;  Conservative   11;  Mismatches   68;  Indels    1;  Gaps    1;

Qy     291 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 350
           |:| :  | |::   :       || | | :||  ||      :   | |        |||
Db       8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 66

Qy     351 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 410
             ||:   ||:  | |     | |        ||||  |:  | | |||||  |  | :  |
Db      67 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 126


RESULT  14
```

```
US-08-700-651-9
; Sequence 9, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 130
;   TYPE: PRT
;   ORGANISM: Cryptosporidium parvum
;   FEATURE:
;   OTHER INFORMATION: mutant/variant of SEQ ID NO:5
US-08-700-651-9


  Query Match             6.6%;  Score 145;  DB 3;  Length 130;
  Best Local Similarity   33.3%;  Pred. No. 2.2e-05;
  Matches   42;  Conservative   12;  Mismatches   62;  Indels   10;  Gaps    2;

Qy     291 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 350
           |:| :  | |::   :       || | | :||  ||    |:   |         |||
Db       8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKPTTTTT---------TTTT 57

Qy     351 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 410
            ||:    ||:  | |     | |     |  |||| |:  |   |||||  |  | : :|
Db      58 TTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTKKPTTTTATTTTTSET 117

Qy     411 VVVQVP 416
           |  |
Db     118 ESVIKP 123


RESULT  15
US-08-928-361B-14
; Sequence 14, Application US/08928361B
  Patent No. 6071518
  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  130 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-14


  Query Match             6.6%;  Score 145;  DB 3;  Length 130;
  Best Local Similarity   33.3%;  Pred. No. 2.2e-05;
  Matches   42;  Conservative   12;  Mismatches   62;  Indels   10;  Gaps    2;

Qy     291 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 350
           |:| :  | |::   :       || | | :||  ||    |:   |         |||
Db       8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKPTTTTT---------TTTT 57

Qy     351 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 410
            ||:    ||:  | |     | |     |  |||| |:  |   |||||  |  | : :|
Db      58 TTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTKKPTTTTATTTTTSET 117

Qy     411 VVVQVP 416
           |  |
Db     118 ESVIKP 123


Search completed: October  3, 2002, 16:30:03
Job time: 151 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:28:17 ; Search time 20.85 Seconds
(without alignments)
1917.178 Million cell updates/sec

Title: US-09-715-876-8_COPY_17_432
Perfect score: 2190
Sequence: 1 AKTITGVFDSFNSLTWSNAA.....TTTTRTNPTDSIDTVVVQVP 416

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2190	100.0	1260	2	S60896	agglutinin-like pr
2	1907	87.1	1419	2	T30531	agglutinin-like ad
3	308.5	14.1	650	2	S22835	alpha-agglutinin -
4	198.5	9.1	948	2	T11678	hypothetical prote
5	177	8.1	1367	2	S51959	hypothetical prote
6	177	8.1	1537	2	S53465	flocculation prote
7	175.5	8.0	1832	2	T31113	mucin-like glycopr
8	174.5	8.0	973	2	T40778	hypothetical 129.5
9	171.5	7.8	456	2	T38221	hypothetical serin
10	171	7.8	1251	2	T21389	hypothetical prote
11	167.5	7.6	662	2	A45155	mucin FIM-C.1 - Af
12	161.5	7.4	1075	2	S48992	flocculation prote
13	161	7.4	1428	2	AC2224	hypothetical prote
14	156.5	7.1	770	2	T22808	hypothetical prote
15	156	7.1	851	2	T22696	hypothetical prote
16	155	7.1	597	2	B75556	hypothetical prote
17	155	7.1	1180	2	E86719	hypothetical prote
18	154.5	7.1	1367	1	S48478	glucan 1,4-alpha-g
19	154	7.0	1283	2	T39174	hypothetical Serin
20	153.5	7.0	725	2	A41258	a-agglutinin core
21	153.5	7.0	3570	2	T45025	mucin MUC5B, trach
22	153	7.0	792	2	S70305	hypothetical prote
23	152	6.9	825	2	T29634	hypothetical prote
24	152	6.9	1902	2	B45764	lactocepin (EC 3.4
25	151.5	6.9	1797	2	T21889	hypothetical prote
26	151.5	6.9	1805	2	T21888	hypothetical prote
27	148.5	6.8	1169	2	S38181	flocculation prote
28	145.5	6.6	648	2	PC4395	mucin 3 - human (f
29	145	6.6	995	2	S50358	hypothetical prote
30	145	6.6	1041	2	S55862	pr
31	145	6.6	1609	2	S25345	prol
32	145	6.6	1962	2	A32634	lactt
33	144.5	6.6	609	2	S62518	hypoth
34	144.5	6.6	881	2	S56032	probab
35	143.5	6.6	3029	2	S76109	hypothe
36	143	6.5	798	2	T34248	hypothet
37	143	6.5	1902	2	S06997	lactocepi C 3.4
38	142.5	6.5	1275	2	T33369	hypothetical prote
39	142	6.5	522	2	S41819	nucleoporin p62 -
40	142	6.5	583	2	S67571	hypothetical prote
41	142	6.5	1777	2	T34369	hypothetical prote
42	141.5	6.5	967	2	S66852	hypothetical prote
43	141	6.4	1902	1	B44858	lactocepin (EC 3.4
44	140.5	6.4	691	2	B75622	hypothetical prote
45	140	6.4	5291	2	F90696	hypothetical prote

ALIGNMENTS

RESULT 1
S60896
agglutinin-like protein - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-2000
C;Accession: S60896
R;Hoyer, L.L.; Scherer, S.; Shatzman, A.R.; Livi, G.P.
Mol. Microbiol. 15, 39-54, 1995
A;Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual
A;Reference number: S60896; MUID:95272392
A;Accession: S60896
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1260 <HOY>
A;Cross-references: EMBL:L25902; NID:g704426; PIDN:AAC41649.1; PID:g704427
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 100.0%; Score 2190; DB 2; Length 1260;
Best Local Similarity 100.0%; Pred. No. 9.2e-136;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432
```

RESULT 2

T30531
agglutinin-like adhesin - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C;Accession: T30531
R;Gaur, N.K.; Klotz, S.A.
Infect. Immun. 65, 5289-5294, 1997
A;Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, tha
A;Reference number: Z20847; MUID:98053977
A;Accession: T30531
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1419 <GAU>
A;Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDN:AAB88883.1
C;Genetics:
A;Gene: ALA1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 87.1%; Score 1907; DB 2; Length 1419;
Best Local Similarity 86.3%; Pred. No. 3.9e-117;
Matches 359; Conservative 21; Mismatches 36; Indels 0; Gaps 0;

```
Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || |||:|:| :|||||||| ||||||||||||||||||||||||||||||||| ||||||||
Db      17 AKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTSANPGDTFILNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           :| || |||||||||||||||||||||||||:| ||||: |:||||| ||||||||||
Db      77 FTASQKSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVNNNLRSSIKALGTVTLPIAFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           |||||| |||||||||||||||||| | :|| | ||||||| | ||  || ||||||:
Db     137 GGTGSSVDLEDSKCFTAGTNTVTFNDGSKKLSIAVNFEKSTVDQSGYLTTSRFMPSLNKI 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           ||:|||||||||||||||:| |||||||||:||||:||:||||:||:|||||||||:|
Db     197 ATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           :| || | |||||||||||||||| :| ||| |:| |||||      | ||||:|||||
Db     257 SSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYK 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTTITTSYVGVTT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432
```

RESULT 3
S22835
alpha-agglutinin - yeast (Saccharomyces cerevisiae)
N;Alternate names: 22K glycoprotein; protein J1418; protein YJR004c
C;Species: Saccharomyces cerevisiae
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 29-Oct-1999
C;Accession: S22835; S51229; A32822; S55192; S57019
R;Hauser, K.; Tanner, W.
FEBS Lett. 255, 290-294, 1989
A;Title: Purification of the inducible alpha-agglutinin of S. cerevisiae and molecular c
A;Reference number: S22835; MUID:90005993
A;Accession: S22835
A;Molecule type: DNA
A;Residues: 1-650 <HAU>
A;Cross-references: EMBL:X16861; NID:g3352; PIDN:CAA34752.1; PID:g3353
A;Accession: S51229
A;Molecule type: protein
A;Residues: 20-24 <HA2>
R;Lipke, P.N.; Wojciechowicz, D.; Kurjan, J.
Mol. Cell. Biol. 9, 3155-3165, 1989
A;Title: AG-alpha-1 is the structural gene for the Saccharomyces cerevisiae alpha-agglut

A;Reference number: A32822; MUID:90014768
A;Accession: A32822
A;Molecule type: DNA
A;Residues: 1-448,'P',450-555,'E',557-580,'L',582-650 <LIP>
A;Cross-references: GB:M28164; NID:g171041; PIDN:AAA34417.1; PID:g171044
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A;Reference number: S55183
A;Accession: S55192
A;Molecule type: DNA
A;Residues: 1-650 <DEH>
A;Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60926.1; PID:g854577
R;de Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56771
A;Accession: S57019
A;Molecule type: DNA
A;Residues: 1-650 <ZAG>
A;Cross-references: EMBL:Z49504; NID:g1015625; PIDN:CAA89526.1; PID:g1015626; MIPS:YJ
C;Genetics:
A;Gene: SGD:SAG1; AGAL1; AGALPHA1
A;Cross-references: SGD:S0003764; MIPS:YJR004c
A;Map position: 10R
C;Keywords: glycoprotein

Query Match 14.1%; Score 308.5; DB 2; Length 650;
Best Local Similarity 25.8%; Pred. No. 9.9e-13;
Matches 116; Conservative 79; Mismatches 190; Indels 65; Gaps 20;

```
Qy      10 SFNSLTWSN------AANYAFKGPGYPTWNAVLGWSL-DGTSANPGDTFTLNMPCVFK-- 60
           : | :|:|| ||  | | | :|: | :| || |||:|| |::
Db      21 NINDITFSNLEITPLTAN---KQPD-QGWTATFDFSIADASSIREGDEFTLSMPHVYRIK 76

Qy      61 -YTTSQTSVDLTADGVKYATCQFYSGE-----EFTTFSTLTCTVNDALKSSIKAFGTVTL 114
             :|||:   ||| :  | : | :    | ||| ||| : | |    |::|
Db      77 LLNSSQTATISLADGTEAFKC-YVSQQAAYLYENTTF---TCTAQNDLSSYNTIDGSITF 132

Qy     115 PIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY---LYAS 171
           : |: ||:   :||::| | :|  |  :  |:  | |    ||:|:  :: |
Db     133 SLNFSDGGSSYEYELENAKFFKSGPMLVKLGNQMSDV---VNF-----DPAAFTENVFHS 184

Qy     172 RVMPSLNKVTTLFVAPQCENGYTSG---TMGFSSSNGDVAIDCSNIHIGITKGLNDWNYP 228
                 :  :  | ||| |  |   : : ||| :| :|||:: :  :   ||| :|
Db     185 GRSTGYGSFESYHLGMYCPNGYFLGGTEKIDYDSSNNNVDLDCSSVQVYSSNDFNDWWFP 244

Qy     229 VSSESFSYTKTCTSNGIQIKY-QNVPAGYRPFIDAYIS-ATDVNQYTLAYTNDYTC---- 282
           |  :  || : : |  : :  |  :::|  |  :||  |  |||
Db     245 QSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNALQSLPANVNTIDHALEFQYTCLDTI 304

Qy     283 AGSRLQSKPFTLR----WTGYKNSDAGSNGIVIVATTRTVTD------STTAVTTLPFNP 332
           | :  ::  | |   : |   | :  :| || :|      || :::|:
Db     305 ANTTYATQFSTTREFIVYQGRNLGTASAKSSFISTTTTDLTSINTSAYSTGSISTV---- 360

Qy     333 SVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTT------- 385
             | ::  : ||:  |   |:  : :|:     : : |   |||:
Db     361 ETGNRTTSEVISHVVTTSTKLSPTATTSLTIAQTSIYSTDSNITVGTDIHTTSEVISDVE 420

Qy     386 TVTSEWTGTITTT-TTRTNPTDSIDTVVVQ 414
           |:: |   |:   |: | | :::| : | |
Db     421 TISRETASTVVAAPTSTTGWTGAMNTYISQ 450
```

RESULT 4
T11678
hypothetical protein SPBC21D10.06c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: T11678
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z17313
A;Accession: T11678

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-948 <SEE>
A;Cross-references: EMBL:AL031536; NID:e1319499; PID:e1319505
A;Experimental source: strain 972h(-)
C;Genetics:
A;Map position: IIR
A;Note: SPBC21D10.06c

  Query Match             9.1%;  Score 198.5;  DB 2;  Length 948;
  Best Local Similarity   23.4%;  Pred. No. 2.6e-05;
  Matches  100;  Conservative   74;  Mismatches  133;  Indels  121;  Gaps   22;

Qy       9 DSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQTSV 68
Db     468 NNFSSISHSSASSLPITPSSYLS-NTTLHSSVQSSQSS---QFTVSVPS---------- 512

Qy      69 DLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGS--- 125
Db     513 --------------STQSYSTSSNFTTPI--TISTSLSSFPTTIVSSSFQYSSLSSNVT 555

Qy     126 STDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFV 185
Db     556 TTNAQSSSLSSSNSSALT------HISSSIVSSGS----SSALSSSTIVSSINSSSSVFI 605

Qy     186 AP-----QCENGY-----TSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 235
Db     606 SSVSSSLQYSSSYVTETTTSGSVGFTTT---IATP-----VGSTAGTVVVDIPTPS---W 654

Qy     236 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 295
Db     655 VTETVTSGSV---------GFTTTIATPVGST--------------AGTVLVDIP-TPS 689

Qy     296 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIP---TTTIT 352
Db     690 W----------------VTETVTSGSVEFTTTIATPVGTTAGTVVVDIPTPSWVTETVT 732

Qy     353 TSYVGVTTSYLTKTAPIGETA-TVIVDVPYHT--TTTVTSEWTGTITTTTTRTNPT-DSI 408
Db     733 SGSVGFTTTIAT---PIGTTAGTVLVDIPTPSWVTETVTS---GSVGFTTTIATPVGTTA 786

Qy     409 DTVVVQVP 416
Db     787 GTVLIDVP 794


RESULT   5
S51959
Hypothetical protein YAL063c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 22-Oct-1999
C;Accession: S51959
R;Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A;Reference number: S51956
A;Accession: S51959
A;Molecule type: DNA
A;Residues: 1-1367 <BUS>
A;Cross-references: EMBL:U12980; GSPDB:GN00001; MIPS:YAL063c
C;Genetics:
A;Gene: MIPS:YAL063c
A;Map position: 1L

  Query Match             8.1%;  Score 177;  DB 2;  Length 1367;
  Best Local Similarity   23.3%;  Pred. No. 0.0011;
  Matches  112;  Conservative   53;  Mismatches  181;  Indels  134;  Gaps   26;

Qy      21 NYAFKGPGYPTWNAVLG-WSLD--GTSANPGDTFTLNM----------PCVFKYTTSQTS 67
Db     104 NWGCKGIGACSNNPIIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDS 162

Qy      68 VDLTADG-VKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAF---------GTV----- 112
Db     163 AILSVGGSIAFECC----AQEQPPITSTNFTIN-----GIKPWNGSPPDNITGTVYMYAG 213

Qy     113 ---------------TLPIAFNV-GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDV 155
Db     214 FYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEG--------YVYTF---DNNLS--- 259

Qy     156 EFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIH 215
Db     260 QPNCTIPDPSNYT-VSTTITTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTA 318

Qy     216 IGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLA 275
Db     319 STIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEP 372

Qy     276 YTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFN 331
Db     373 WNSTFTSTSTEL------------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWN 419

Qy     332 PSVDKTKT-----------------IEILQPIPTTTITT-------SYVGVTTSYLTKTA 367
Db     420 DTFTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTG 479

Qy     368 ----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVV 413
Db     480 TNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTD--ETIIV 536


RESULT   6
S53465
flocculation protein FLO1 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YAR050w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C;Accession: S53465; S43543; S57851; S31230; S57349
R;Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Ka
submitted to the EMBL Data Library, February 1994
A;Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of th
A;Reference number: S53458
A;Accession: S53465
A;Molecule type: DNA
A;Residues: 1-1537 <BUS>
A;Cross-references: EMBL:L28920; NID:g1616966; PIDN:AAC09499.1; PID:g694125; MIPS:YAR
R;Watari, J.; Takata, Y.; Ogawa, M.; Sahara, H.; Koshino, S.; Onnela, M.L.; Airaksine
Yeast 10, 211-225, 1994
A;Title: Molecular cloning and analysis of the yeast flocculation gene FLO1.
A;Reference number: S43543; MUID:94262325
A;Accession: S43543
A;Molecule type: DNA
A;Residues: 1-428,'M',430-473,'M',475-518,'M',520-549,'T',551-608,'L',610-636,'M',638
A;Cross-references: EMBL:X78160
R;Onnela, M.
submitted to the EMBL Data Library, September 1994
A;Reference number: S57851
A;Accession: S57851
A;Molecule type: DNA
A;Residues: 1-428,'M',430-463,'D',465-473,'M',475-518,'M',520-549,'T',551-608,'L',610
A;Cross-references: EMBL:X78160; NID:g535933; PIDN:CAA55024.1; PID:g535934
R;Teunissen, A.W.R.H.; Holub, E.; van der Hucht, J.; van den Berg, J.A.; Steensma, H
Yeast 9, 423-427, 1993
A;Title: Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerev
A;Reference number: S31230; MUID:93289821
A;Accession: S31230
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-296,927-1516,'TAYWPVVV' <TEU>
R;Bidard, F.; Bony, M.; Blondin, B.; Dequin, S.; Barre, P.
Yeast 11, 809-822, 1995
A;Title: The Saccharomyces cerevisiae FLO1 flocculation gene encodes for a cell surfa
A;Reference number: S57349; MUID:96090130
A;Accession: S57349
```

```
A;Molecule type: DNA
A;Residues: 1243-1274;1308-1339;1359-1390 <BID>
C;Genetics:
A;Gene: SGD:FLO1
A;Cross-references: SGD:S0000084; MIPS:YAR050w
A;Map position: 1R
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1537/Product: flocculation protein FLO1 #status predicted <MAT>
F;953-997/Domain: repeat A2 <RA2>
F;998-1042/Domain: repeat A3 <RA3>
F;1043-1081/Domain: repeat A4 <RA4>
F;1226-1276/Domain: repeat B1 <RB1>
F;1277-1284/Domain: repeat B2 (partial) #status atypical <RB2>
F;1291-1341/Domain: repeat B3 <RB3>
F;1342-1392/Domain: repeat B4 <RB4>
F;1408-1416/Domain: repeat C1 <RC1>
F;1417-1425/Domain: repeat C2 <RC2>
  1426-1434/Domain: repeat C3 <RC3>
  135,187,262,1114/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             8.1%;  Score 177;  DB 2;  Length 1537;
  Best Local Similarity   23.0%;  Pred. No. 0.0013;
  Matches  112;  Conservative   47;  Mismatches  181;  Indels  146;  Gaps   25;

Qy        21 NYAFKGPG-YPTWNAVLGWSLD--GTSANPGDTFTLNM---------PCVFKYTTSQ-- 65
Db       104 NWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDS 162

Qy        66 --------------------TSVDLTADGVK----------YATCQFYSGEEFTTFSTL 94
Db       163 AILSVGGATAFNCCAQQQPPITSTNFTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---- 218

Qy        95 TCTVNDALKSSIKAFGT----VTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKD 150
Db       219 -----KVVYSNAVSWGTLPISVTLP-----DGTTVSDDFEG--------YVYSFDD---- 256

Qy       151 ISIDVEFEKSTV-DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI 209
Db       257 ---DLSQSNCTVPDPSNYA-VSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVI 312

Qy       210 DCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDV 269
Db       313 RTPTTASTIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETII------VIRTPTTATTA 366

Qy       270 NQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVT 326
Db       367 ITTTEPWNSTFTSTSTEL-------------TTVTGTNGLPTDETIIVIRTPTTATTAMT 413

Qy       327 TL-PFNPSVDKTKT-----------------IEILQPIPTTTITT-------SYVGVTTS 361
Db       414 TTQPWNDTFTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTE 473

Qy       362 YLTKTA----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTTTRTN--PTDS 407
Db       474 LTTVTGTNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTD- 531

Qy       408 IDTVVV 413
Db       532 -ETIIV 536

RESULT   7
T31113
mucin-like glycoprotein 900 - Cryptosporidium parvum
C;Species: Cryptosporidium parvum
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31113
R;Barnes, D.A.; Bonnin, A.; Huang, J.X.; Gousset, L.; Wu, J.; Gut, J.; Doyle, P.; Dubrem
Mol. Biochem. Parasitol. 96, 93-110, 1998
A;Title: A novel multi-domain mucin-like glycoprotein of Cryptosporidium parvum mediates
A;Reference number: Z20989; MUID:99066935
A;Accession: T31113
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1832 <BAR>
A;Cross-references: EMBL:AF068065; NID:g4063041; PID:g4063042; PIDN:AAC98153.1

  Query Match             8.0%;  Score 175.5;  DB 2;  Length 1832;
  Best Local Similarity   22.9%;  Pred. No. 0.002;
  Matches   93;  Conservative   45;  Mismatches  176;  Indels   93;  Gaps   13;

Qy        43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
Db       447 TTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTKKPTTTTTTTTTTTTKK 506

Qy       103 KSSIKAFGTVTLPIAFNVGGTGSSTDL-----------EDSKCFTAGTNTVTFNDGDKDI 151
Db       507 PTTTTTATTTT---------TTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRI 557

Qy       152 SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI-- 209
Db       558 ENGMAFTMIPNDDTHVRFRFKVKDVGNTISV-----RCGKG--AGKLEFPDRSLDFTIPP 610

Qy       210 -----DCSNIHIGITKGLNDWNYPVSSESFS---------------YTKTCTSNGIQIKY 249
Db       611 VAGHNSCS-IIVGVSGGGKIHVSPYGSKDVSLISAPIQPCELFNEVYCDTCTA-----KY 664

Qy       250 QNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI 309
Db       665 GAIHSGYQTSADFVTTTT---------------------AKPTT---------TTTGAPGQ 695

Qy       310 VIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPI 369
Db       696 PTTTTTGSPSKPTTTTTT-------KATTTTTILNPIITTT-TQKPTTTTTTKVPGKPPI 747

Qy       370 GETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
Db       748 ATTTTTLKPIVTTTTTKATTTTTTTVPTTTT-TTKRDEMTTTTTPLP 793

RESULT   8
T40778
hypothetical 129.5 kd protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T40778
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21884
A;Accession: T40778
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-973 <LYN>
A;Cross-references: EMBL:AL021837; PIDN:CAA17032.1; GSPDB:GN00067; SPDB:SPBC947.04
A;Experimental source: strain 972h-; cosmid c947
C;Genetics:
A;Gene: SPDB:SPBC947.04
A;Map position: 2

  Query Match             8.0%;  Score 174.5;  DB 2;  Length 973;
  Best Local Similarity   22.6%;  Pred. No. 0.001;
  Matches  112;  Conservative   63;  Mismatches  157;  Indels  163;  Gaps   25;

Qy        15 TWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQTSV-DLTAD 73
Db       149 TWSTNL------PTNPTTTAI--YSTSGSS-------NITTPYSNRITNSNTSVNDITSK 193

Qy        74 --GVKYATCQFYSGEEF---------TTFSTLTCTV-------------NDALKSSIKAF 109
Db       194 YLSVGTITLTTISGSDLYTSTFPANGTTSGTVEVVIPTAGTVTETAVSGSELYTSTFPAN 253

Qy       110 GTV--TLPIAFNVGGTGSSTDLEDSKCF-------------------TAGTNTVTFNDGD 148
```

```
Db    254 GTTSGTVEVVIPTAGTRTVTKISGSKFFTTTTDASGTVSGTVEVVLPTAGTNTMTVVSGS 313

Qy    149 K------DISIDVEFEKSTVDPSAYLYASR----------------------VMPSLN 178
Db    314 RFFTSVVSASGTVSGEQVIVYPTAGMVTETIVSGSEIFNTTYPASGTRTGTVEVVIPTAG 373

Qy    179 KVT-------TLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSS 231
Db    374 TVTETEISGSELYTSTFPANGTTSGTV-------EVVIPTAGTRT-VTK--------ISG 417

Qy    232 ESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYT-CAGSRLQSK 290
Db    418 SKF-FTTTTDASGT-----------------VSGTVEVVLPTAGTNTMTVVSGSRF--- 455

Qy    291 PFTLRWTGYKNSDAGSNG--IVIVATTRTVTDSTTA-----VTTLPFNPSVDKTKTIEIL 343
Db    456 -----FTSVVSASGTVSGEHIIVEPTAGVVTETVVSGSVGYTTTYPAHDTV--SGTVEVV 508

Qy    344 QP---IPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHT----TTTVTSEWTGTIT 396
Db    509 EPTAGVVTETVVSGSVGYTTAY-----PAHDTVSGTVEVVEPTAGVVTETVVSGSVGYTT 563

Qy    397 TTTTRTNPTDSIDTV 411
Db    564 TYPAHDTVSGTVEVV 578


RESULT   9
T38221
hypothetical serine-threonine rich protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 17-Nov-2000
C;Accession: T38221
R;Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21780
A;Accession: T38221
A;Molecule type: DNA
A;Residues: 1-456 <MUR>
A;Cross-references: EMBL:AL021813; PIDN:CAA16975.1; GSPDB:GN00066; SPDB:SPAC23A1.01c
A;Experimental source: strain 972h-; cosmid c23A1
C;Genetics:
A;Gene: SPDB:SPAC23A1.01c
A;Map position: 1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

  Query Match             7.8%;  Score 171.5;  DB 2;  Length 456;
  Best Local Similarity   24.8%;  Pred. No. 0.0006;
  Matches  110;  Conservative   57;  Mismatches  147;  Indels  129;  Gaps   24;

Qy     10 SFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQ---- 65
Db     94 SFTSIR-DDSVIYAQSGTFY-----IVGG--EGISSTTGSTF----QSMTTFTSSQTNSG 141

Qy     66 ----------TSVDLTADGVKY--ATCQF-YSGEEFTTFSTLTCTVNDAL-----KSSIK 107
Db    142 HASASTSIPSTAITVTANSTIYSSATSSFPYS----TDVSVSTGTSTDIVTLPPPASSTS 197

Qy    108 AFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 167
Db    198 SFSTIT-----------NTSMIPSSSSFTTTTGSPYYNTS--------SFLPSSVISSAS 238

Qy    168 LYASRVMP----------------SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC 211
Db    239 LSSSSVLPTSIITSTSTPVTVSSSSLSSFTPSYSTNLTTTGSTTTTGSATVSSSPFYSNS 298

Qy    212 SNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQ 271
Db    299 SVIPTSVPSSVSSF----TSSSSSYTTTLTASNTSVTYTGTGTG---------SAT---- 341

Qy    272 YTLA---YTNDY---TCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAV 325
Db    342 FTSSPPFYSNSSVIPTSVPSSVSS--FTSSNSSYTTTLTASN--TSITYTGTGTGSATFT 397

Qy    326 TTLPF--NPSVDKTKTIEILQPIPTT--TITTSYVGVTTSYLTKTAPIGETATVIVDVPY 381
Db    398 SSPPFYSNSSV----------IPTSVPSSVSSFTSSNSSYTT-------TLTA------ 433

Qy    382 HTTTTVTSEWTGTITTTTTRTNP 404
Db    434 -SNTTVTFTGTGTGSATFTSSPP 455


RESULT  10
T21389
hypothetical protein F26C11.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21389
R;Matthews, P.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19416
A;Accession: T21389
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1251 <WIL>
A;Cross-references: EMBL:Z47072; PIDN:CAA87369.1; GSPDB:GN00020; CESP:F26C11.3
A;Experimental source: clone F26C11
C;Genetics:
A;Gene: CESP:F26C11.3
A;Map position: 2
A;Introns: 24/1; 111/1; 208/1; 272/1; 380/1; 394/3; 485/3; 586/3; 630/3; 669/3; 713/3

  Query Match             7.8%;  Score 171;  DB 2;  Length 1251;
  Best Local Similarity   23.8%;  Pred. No. 0.0024;
  Matches  107;  Conservative   60;  Mismatches  154;  Indels  128;  Gaps   23;

Qy     27 PGYPTWNAVLGWSLDGTS----------------ANPGDTFTLNMPCVFKYTTSQTSVDLT 71
Db    609 PGTTTYN----WPTGGTTRMLPSGEIILSESLIAYPNCT-TVLMQLIYTPSTNKTRTETT 663

Qy     72 AD--GVKYATCQFYSGEEFT-----TFSTLTCTVN--------------DALKSSIKAF- 109
Db    664 TDTEGCKKTSTISSSSSKFSITPTPTPSSGTTTYNWPTGGTTRTLPSGEIILSESLIAFQ 723

Qy    110 --GTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 167
Db    724 NCTTVLMQLIYN-----PSTN-------KTRTETTTDAEGCKKTSSTSKISTTPTSPT-- 769

Qy    168 LYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNY 227
Db    770 --SSKPTPTSTSMTTTYNWP---TGGTTRTL----PSGEI-------------------- 800

Qy    228 PVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN--QYTLAYTNDYTCAGS 285
Db    801 -ILSESLIAYKNCTTVLMQLIYN--PSTNKTRTE---TTTDAQGCKATSSTSLKPTSPSS 854

Qy    286 RLQSKP-FTLRW-TGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKT---- 339
Db    855 STASPPTTTYNWPTGGTTRTLPSGEIILSESLIAYKNCTTVLMQLIYNPSTNKTRTETTT 914

Qy    340 ------IEILQPIPTTTITTSYVGVTTSYLTKTAPIGE-----------TATVIVDVPY 381
Db    915 DAQGCKATITTPTP---ITTTY-NWPTGGTTRTLPSGEIILSESLIAYKNCTTVLMQLIY 970

Qy    382 H-----TTTTVTSEWTGTITTTTTRTNPT 405
Db    971 NPSTNKTRTETTSDAQGCKATSTTQTPTT 999


RESULT  11
A45155
mucin FIM-C.1 - African clawed frog (fragment)
```

```
C;Species: Xenopus laevis (African clawed frog)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 21-Jul-2000
C;Accession: A45155
R;Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A;Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1
A;Reference number: A45155; MUID:93077556
A;Accession: A45155
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-662 <HAU>
A;Cross-references: GB:L02115; NID:g214147; PIDN:AAA74725.1; PID:g951460
C;Superfamily: trefoil homology
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>

  Query Match             7.6%;  Score 167.5;  DB 2;  Length 662;
  Best Local Similarity   24.3%;  Pred. No. 0.0018;
  Matches   94;  Conservative   28;  Mismatches  166;  Indels   99;  Gaps   16;

Qy      42 GTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDA 101
           | |     | |   |    :  |: |:   |    |             |  :| |
Db     206 GHSHEEHTTTTTKAPTTIQIATTTTTPTTTTTTTK----------ATPTTTTTTKATPT 254

Qy     102 LKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKST 161
             :: ||  | | |          :|    :|  |  | |                  :|
Db     255 TTTTTKATTTTTTP----------TTTTTTTKATTTPTTT-----------------TT 286

Qy     162 VDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKG 221
             |:       :  | ||          |||      |  :   ||   : |||:
Db     287 TTPTT---------TTTKATTTTTT-------TSGECKMEPSKRE---DCG--YSGITES 325

Qy     222 LNDWNYPVSSESPSYTKTCTSNGIQI------KYQNVPAGYRPFIDAYISATDVNQYTLA 275
                     |   || |     |:        | |  |:|      |:|    |
Db     326 QCRTKGCCFDSSIPQTKWCFYTLSQVADCKVEPSQRVDCGFRG-----ITADQCRQKNCC 380

Qy     276 YTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVD 335
           :  | : :|:         :|  |  |          || |:| |  ||  || |   :
Db     381 F--DSSISGT---------KWCFYSTSQ--------VAATKTTTPTT--TTTPTTTTTT 419

Qy     336 K----TKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEW 391
           |    | |       | ||| ||:    ||:  | | |    |        |  |||| |:
       420 KATTTTPTTTTTTPTTTTTTTTT-TKATTTTPTTTTPTTTTTKATTTTPTTTTTTPTTTT 478

Qy     392 TGTITT--TTTRTNPTDSIDTVVVQVP 416
           |   ||  ||| | || :        |
Db     479 TKATTTTPTTTTTTPTTTTTKATTTTP 505


RESULT  12
S48992
flocculation protein homolog YHR211w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 29-Oct-1999
C;Accession: S48992
R;Macri, C.
submitted to the EMBL Data Library, February 1994
A;Description: The sequence of S. cerevisiae cosmid 9177.
A;Reference number: S46671
A;Accession: S48992
A;Molecule type: DNA
A;Residues: 1-1075 <MAC>
A;Cross-references: EMBL:U00029; NID:g551322; PID:g458919; GSPDB:GN00008; MIPS:YHR211w
C;Genetics:
A;Gene: MIPS:YHR211w
A;Map position: 8R

  Query Match             7.4%;  Score 161.5;  DB 2;  Length 1075;
  Best Local Similarity   22.2%;  Pred. No. 0.0083;
  Matches  110;  Conservative   57;  Mismatches  209;  Indels  119;  Gaps   22;

Qy      21 NYAFKGPG-YPTWNAVLGWSLD--GTSANPGDTFTLNM------PCVFKYTTSQTSVDLT 71
           |:  || |       :  || |  |    | :  || |       |    || |  :|| :
Db     104 NWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFSFATVDDS 162

Qy      72 A-----DGVKYATCQFYSGEEFTTFSTLTCTVN-------------------------- 99
           |     : :  |     :|    ::     |:|
Db     163 AILSVGGSIAFECC----AQEQPPITSTNFTINGIKPWDGSLPDNITGTVYMYAGYYYPL 218

Qy     100 DALKSSIKAFGTVTLPIAFNVGGTGSST----------DLEDSKC-------FTAGTNTV 142
             : |:  ::||: : :    | | |            ||  | |         |  | |
Db     219 KVVYSNAVSWGTLPISVELPDGTTVSDNFEGYVYSFDDDLSQSNCTIPDPSIHTTSTITT 278

Qy     143 -------TFNDGDKDISIDVEFEKSTVDPSAYLY-----ASRVMPSLNKVTTLFVAPQCE 190
                 ||     :::   :       | :  :      || :  :    |  | :   |
Db     279 TTEPWTGTFTSTSTEMTTITDTNGQLTDETVIVIRTPTTASTITTTTEPWTGTFTSTSTE 338

Qy     191 NGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQ 250
             :|| |  :    : |        ||   | |    :| |    |     :|| |   :
Db     339 MTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNG-QPTDE 397

Qy     251 NVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSD------- 303
           |     |  :  |: |      |  :|  :|   : :     |:  |  : :|
Db     398 TVIVIRTPTSEGLITTT-----TEPWTGTFTSTSTEVT----TITGTNGQPTDETVIVIR 448

Qy     304 -AGSNGIVIVAT---TRTVTDSTTAVTTLPFNPSVDKTKTIEILQP------IPTT---- 349
            | |::   |   | | | ::| :||:     |   :|: :::     |  ||
Db     449 TPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPW 508

Qy     350 --TITTSYVGVTTSYLTKTAPIGETATVI-VDVPYHTTTTVTSEWTGTIT------TTTT 400
             | |::   |||   |   |  ||  ||         || |  |||| |      || |
Db     509 TGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVT 568

Qy     401 RTN--PTDSIDTVVV 413
           ||  |||  :||:|
Db     569 GTNGQPTD--ETVIV 581


RESULT  13
AC2224
hypothetical protein all3346 [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AC2224
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2224
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1428 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75045.1; PID:g17132441; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all3346


  Query Match             7.4%;  Score 161;  DB 2;  Length 1428;
  Best Local Similarity   24.5%;  Pred. No. 0.013;
  Matches  118;  Conservative   43;  Mismatches  191;  Indels  130;  Gaps   24;

Qy      13 SLTWSNAAN--YAFKGPGYPTWNAVLGWSLDGTSANPGD-------------TFTLNMP 56
           |:|     |  | |   |  |    : :|: ||:    |             | |  :
Db     808 SVTEDGTPNLIYTFTRTGSTTNALTVNYSVAGTATLNTDYAQTGAASFTATTGTITFAVG 867

Qy      57 CVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCT---VNDALKS--------S 105
                 |   : |.| :  :       || |:|  :|   |    ||   |         |
```

```
Db    868 ASTAILTINPTADTTVESNETVALTLASGTGYTVGTTTAVTGTITNDDFPSITLAVSPAS 927

Qy    106 IKAFGTV-------------TLPIAFNVGGTGS-STDLEDS--KCFTAGTNTVTFNDGD 148
Db    928 VTEDGTPNLIYTFTRTGSTTNALTINFGVAGTATLNTDYAQSGAASFTATTGTITFAAGA 987

Qy    149 KDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVA 208
Db    988 STAIL-------TINPTA-----DTTVESNETVALTLA--SGTGYTVGTT--TAVTGTIT 1031

Qy    209 IDCSNIHIGITKGLNDWNYPV----SSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI 264
Db   1032 ND-DTLPTGITINLSGSQTIVEGNSSPQNVTYTVTLSQASSQI-----------ITVQYA 1079

Qy    265 SATDVNQYTLAYTNDYTCAGSRLQSKP---------------------FTLRWTGYKNS 302
Db   1080 TANG----TATAGSDYTSTTGTLTFNPGETSKVINIPILNDSVNEANETFTLRLTSPTNA 1135

Qy    303 DAGSNGIVIVATTRTVTD--STTAVTTLPFNPSVDKTKTIEILQPIPTTTIT-TSYVGVT 359
Db   1136 TLGTTNTV----TTTITDTLSASVTTTLPTN--------VENLTLTGTAAINGTGNAG-- 1181

Qy    360 TSYLTKTAPIGETATVIVD--VPYHTTTTVTSEWTGTITTTTTTRTNPTDSID----TVVV 413
Db   1182 NNILT-----GNSGNNILSGGAGNDTYAFVANAALGTDTITETATGGIDTIDFNGSTATV 1236

Qy    414 QV 415
Db   1237 RV 1238
```

RESULT 14
T22808
hypothetical protein F56H9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22808
R;Burton, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19618
A;Accession: T22808
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-770 <WIL>
A;Cross-references: EMBL:Z74473; PIDN:CAA98949.1; GSPDB:GN00023; CESP:F56H9.1
A;Experimental source: clone F56H9
C;Genetics:
A;Gene: CESP:F56H9.1
A;Map position: 5
A;Introns: 235/1; 262/2; 320/1; 367/2; 510/3; 654/1; 681/2

Query Match 7.1%; Score 156.5; DB 2; Length 770;
Best Local Similarity 20.3%; Pred. No. 0.011;
Matches 106; Conservative 48; Mismatches 170; Indels 197; Gaps 17;

```
Qy     50 TFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAF 109
Db    166 TTTTTVPATTTSTATTTTVPPTTSTTTTTV-----PPTTTSTTTTTTVPPTTTSSTTTT 220

Qy    110 GTVTLP-----------------------------------IAFNVGGT---GSSTDL 129
Db    221 TTTTLPTTTTTMMIPICSLDSLTFGKGDNFQPQVDIDVTYSNLVANIIPGTMDMQSLSTM 280

Qy    130 EDSKCFTAGTN------------------TVTFNDGDKDISID-------VEFEKST-- 161
Db    281 QISCTAIAGYNAYMIFDVFNQPAENQGFPQSVTVNAQCNRILINLFMFFKFLSYEKTSFY 340

Qy    162 --------------------VDPSAYLYASRVMPSLN--------------KVTTLFV- 185
Db    341 ILCTSKTLSNFILLTVYFLYIGPTELLYTQIGSMNLNTYLNQTMGLGMYLQGPITQMIIT 400

Qy    186 ------------APQCENGYTSGTM----------------GFSSSNG-DVAIDCSNI 214
```

```
Db    401 INRFLVIWFTPTHVPQYSHRITLGALSVSWITVTWLSTLIGLPGFDNNNNFDIIINLANC 460

Qy    215 HIGITKGLNDWNYPVSSESFSYTKTCTSNGIQI-------------KYQNVPAGYRPF-- 259
Db    461 RV-----------PIGFEHIGYYSTPCNNQITIIVVSGIFLIGFLTNFMNFMIGGKLIYT 509

Qy    260 -----IDAYIS---------ATDVNQYTLAYTNDY--TCAGSRLQSKPFTLRWTGYKNSD 303
Db    510 WKTLTIFRFLSNFKCVMKWISTLLHILLLTYKLDVINSCIATSPTADPTT-------TTT 562

Qy    304 AGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYL 363
Db    563 EATTTTEITTTTEEVTTTTEPVTTTTTPTTTSTTSTTTTTTPTTTTTTTPTTTTSTTSTT 622

Qy    364 TKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTTRTNP 404
Db    623 TTTTPTTTTST--------TTTTTTTTTTATPTTTTTTMPP 655
```

RESULT 15
T22696
hypothetical protein F55B11.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22696
R;Ainscough, R.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19601
A;Accession: T22696
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-851 <WIL>
A;Cross-references: EMBL:Z83318; PIDN:CAB05903.1; GSPDB:GN00022; CESP:F55B11.3
A;Experimental source: clone F55B11
C;Genetics:
A;Gene: CESP:F55B11.3
A;Map position: 4
A;Introns: 49/3; 123/3; 226/1; 282/3; 669/3; 743/3

Query Match 7.1%; Score 156; DB 2; Length 851;
Best Local Similarity 23.0%; Pred. No. 0.014;
Matches 87; Conservative 32; Mismatches 170; Indels 90; Gaps 11;

```
Qy     43 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 102
Db    330 TTPSELSTTTASVP------TTTTSVPTTTTTVPTTT---------TTVPTTTTTVSTT- 373

Qy    103 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 162
Db    374 --------TTTVP---------TTTTTVPTTTTTVSTTTTT----------VPTTTTTV 405

Qy    163 DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGL 222
Db    406 P----------------TTTTTVPTTTTTVPTTTTSVPTTTTTVPTTTTTV-------- 440

Qy    223 NDWNYPVSSESFSYTKT---CTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTND 279
Db    441 -----PTTTTTVSTTTTTVPTTTTTVSTTTTTVPT---------TTTTVPTTTTTVPTT 485

Qy    280 YTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKT 339
Db    486 TTTVPTTTTSVPTTTTTVPTTTTTVPTTTTTVSTTTTVPTTTTTVPTTTTTVSTTTTTV 545

Qy    340 IEILQPIPTTTITTSYVGVTTSYLTKTAPIGETA--TVIVDVPYHTTTTVTSEWTGTITT 397
Db    546 PTTTTTVPTTTTTVSTTTTTVPTTTTTAPTTTTTVPTTTTTVP-TTTTTVSTTTTTVPTT 604

Qy    398 TTTRTNPTDSIDTVVVQVP 416
Db    605 TTTVPTTTTTVPTTTTTVP 623
```

Search completed: October 3, 2002, 16:30:32
Job time: 135 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:30:07 ; Search time 13.4 Seconds
(without alignments)
1202.040 Million cell updates/sec

Title: US-09-715-876-8_COPY_17_432
Perfect score: 2190
Sequence: 1 AKTITGVFDSFNSLTWSNAA.....TTTTRTNPTDSIDTVVVQVP 416

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2178	99.5	1260	1	ALS1_CANAL	P46590 candida alb
2	1907	87.1	1419	1	ALA1_CANAL	O13368 candida alb
3	1889	86.3	1119	1	ALS3_CANAL	O74623 candida alb
4	1773.5	81.0	468	1	ALS2_CANAL	O74657 candida alb
5	1600	73.1	469	1	ALS4_CANAL	O74660 candida alb
6	308.5	14.1	650	1	SAG1_YEAST	P20840 saccharomyc
7	177	8.1	1322	1	YAG3_YEAST	P39712 saccharomyc
8	177	8.1	1537	1	FLO1_YEAST	P32768 saccharomyc
9	171	7.8	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
10	167.5	7.6	662	1	MUC1_XENLA	Q05049 xenopus lae
11	161.5	7.4	1075	1	FLO5_YEAST	P38894 saccharomyc
12	160	7.3	567	1	CHI3_CANAL	P40954 candida alb
13	156	7.1	540	1	GUX1_ASPAC	O59843 aspergillus
14	154.5	7.1	1367	1	AMYH_YEAST	P08640 saccharomyc
15	153.5	7.0	725	1	AGA1_YEAST	P32323 saccharomyc
16	152	6.9	1902	1	P1P_LACLC	P16271 lactococcus
17	148.5	6.8	1169	1	YK82_YEAST	P36170 saccharomyc
18	145	6.6	995	1	YIQ9_YEAST	P40442 saccharomyc
19	145	6.6	1041	1	EGT2_YEAST	P42835 saccharomyc
20	145	6.6	1609	1	FIG2_YEAST	P25653 saccharomyc
21	145	6.6	1902	1	P3P_LACLC	P15292 lactococcus
22	144.5	6.6	881	1	YJH8_YEAST	P47033 saccharomyc
23	143	6.5	1902	1	P2P_LACLC	P15293 lactococcus
24	142	6.5	522	1	NU62_HUMAN	P37198 homo sapien
25	141	6.4	1902	1	P2P_LACPA	Q02470 lactobacill
26	138	6.3	605	1	YHC8_YEAST	P38739 saccharomyc
27	135.5	6.2	400	1	MUA1_XENLA	P10667 xenopus lae
28	135	6.2	670	1	VG50_HSV11	Q00130 ictalurid h
29	133	6.1	827	1	CSG_HALVO	P25062 halobacteri
30	132.5	6.1	562	1	CHIT_YEAST	P29029 saccharomyc
31	132.5	6.1	796	1	YS8A_CAEEL	Q09625 caenorhabdi
32	131.5	6.0	1306	1	MSB2_YEAST	P32334 saccharomyc
33	130.5	6.0	338	1	SED1_YEAST	Q01589 saccharomyc
34	130	5.9	525	1	NU62_RAT	P17955 rattus norv
35	128	5.8	907	1	VGP3_EBV	P03200 epstein-bar
36	128	5.8	5179	1	MUC2_HUMAN	Q02817 homo sapien
37	127.5	5.8	598	1	YAUF_SCHPO	Q10168 schizosacch
38	127.5	5.8	886	1	VGP3_EBVA8	Q07284 epstein-bar
39	127	5.8	600	1	SP96_DICDI	P14328 dictyosteli
40	126.5	5.8	1802	1	HKR1_YEAST	P41809 saccharomyc
41	126	5.8	445	1	WAPA_STRMU	P11000 streptococc
42	125	5.7	767	1	AMYH_SACDI	P04065 saccharomyc
43	125	5.7	1140	1	YM96_YEAST	Q04893 saccharomyc
44	125	5.7	1161	1	DAN4_YEAST	P47179 saccharomyc
45	124.5	5.7	768	1	AMYI_SACDI	P29760 saccharomyc

ALIGNMENTS

```
RESULT   1
ALS1_CANAL
ID   ALS1_CANAL     STANDARD;      PRT;  1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST SAG1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L25902; AAC41649.2; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      486    491       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
```

```
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-THR.
FT   DOMAIN      874    877       POLY-SER.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;


  Query Match             99.5%;  Score 2178;  DB 1;  Length 1260;
  Best Local Similarity   99.5%;  Pred. No. 4.8e-140;
  Matches  414;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     257 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRSQSKPFTLRWTGYK 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432


RESULT   2
ALA1_CANAL
ID   ALA1_CANAL     STANDARD;      PRT;  1419 AA.
AC   O13368;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein ALA1 precursor (Agglutinin-like adhesin).
GN   ALA1 OR ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98053977; PubMed=9393828;
RA   Gaur N.K., Klotz S.A.;
RT   "Expression, cloning, and characterization of a Candida albicans
RT   gene, ALA1, that confers adherence properties upon Saccharomyces
RT   cerevisiae for extracellular matrix proteins.";
RL   Infect. Immun. 65:5289-5294(1997).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF025429; AAB88883.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1419       AGGLUTININ-LIKE PROTEIN ALA1.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      437    441       POLY-THR.
FT   DOMAIN      673    676       POLY-SER.
FT   DOMAIN      687    690       POLY-SER.
FT   DOMAIN      700    703       POLY-SER.
FT   DOMAIN      719    724       POLY-SER.
FT   DOMAIN      749    752       POLY-SER.
FT   DOMAIN      787    791       POLY-SER.
FT   DOMAIN      869    872       POLY-SER.
FT   DOMAIN      875    883       POLY-SER.
FT   DOMAIN      901    911       POLY-SER.
FT   DOMAIN     1216   1221       POLY-SER.
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1301   1301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1326   1326       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1419 AA;  149635 MW;  249F33F688A9D5B6 CRC64;


  Query Match             87.1%;  Score 1907;  DB 1;  Length 1419;
  Best Local Similarity   86.3%;  Pred. No. 1.3e-121;
  Matches  359;  Conservative   21;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || |||:|:| :|||||| ||||||||||||||||||||||||||||| ||||||||
Db      17 AKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTSANPGDTFILNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           :| || |||||||||||||||||||||||:| ||||: |:||||| ||||||||||
Db      77 FTASQKSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVNNNLRSSIKALGTVTLPIAFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           |||||| |||||||||||||||| | :|| | ||||||| | ||  || ||||||:
Db     137 GGTGSSVDLEDSKCFTAGTNTVTFNDGSKKLSIAVNFEKSTVDQSGYLTTSRFMPSLNKI 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           ||:|||||||||||||:| |||||||||:||||:||:||||:||:||||||||:|
Db     197 ATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           :| || | |||||||||||||||| :| ||| |:| |||||      | ||||:|||||
Db     257 SSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYK 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432


RESULT   3
ALS3_CANAL
```

```
ID   ALS3_CANAL     STANDARD;      PRT;  1119 AA.
AC   O74623;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 3 precursor.
GN   ALS3.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98309840; PubMed=9644209;
RA   Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;
RT   "Candida albicans ALS3 and insights into the nature of the ALS gene
RT   family.";
RL   Curr. Genet. 33:451-459(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U87956; AAC39486.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1119       AGGLUTININ-LIKE PROTEIN 3.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      557    563       POLY-THR.
FT   DOMAIN      593    597       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    777       POLY-THR.
FT   DOMAIN     1044   1047       POLY-THR.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    759    759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    845    845       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    987    987       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1050   1050       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1119 AA;  119927 MW;  6A3FB3FC8C879A71 CRC64;

  Query Match             86.3%;  Score 1889;  DB 1;  Length 1119;
  Best Local Similarity   84.9%;  Pred. No. 1.6e-120;
  Matches  353;  Conservative   30;  Mismatches   33;  Indels    0;  Gaps    0;
```

```
Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           ||||||||:||||||||||| | :|||| ||||||||||||||||:||||||||||||
Db      17 AKTITGVFNSFNSLTWSNAATYNYKGPGTPTWNAVLGWSLDGTSASPGDTFTLNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           :|||||||||| ||||||||| :|||| |||||||||:: |  |||| ||||||:||||
Db      77 FTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           |||||| |||||||||||||||||| | |||:|:||:| |||  ||  |||:||||||
Db     137 GGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKV 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           :|||||||| ||||||||||::: ||| |||||||:||||||||||||||||||||||||
Db     197 STLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           :|||| | |:|||||||||:|||||||||| |||:| |:|||||   |  |||||||||:
Db     257 SSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYR 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           |||||||||||||||||||||||||||:|: |||||||||:||||||||||||||||||
Db     317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || |||||||||||||||:||||||||||:||||||:||| ||||||||||:||||
Db     377 SYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVP 432
```

```
RESULT   4
ALS2_CANAL
ID   ALS2_CANAL     STANDARD;      PRT;   468 AA.
AC   O74657;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 2 precursor (Fragment).
GN   ALS2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF024580; AAC64235.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   >468       AGGLUTININ-LIKE PROTEIN 2.
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50203 MW;  BFE773E169ED0FAF CRC64;

  Query Match             81.0%;  Score 1773.5;  DB 1;  Length 468;
  Best Local Similarity   80.3%;  Pred. No. 3.6e-113;
```

```
  Matches  334;  Conservative   27;  Mismatches   54;  Indels    1;  Gaps    1;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || ||||||:||:|||||: | |||:|||  |||||||||||||||||||||||||||||||
Db      17 AKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : | ||||||||:|||||||||||||||||||:| |||:: | ||||| |||||||:|||
Db      77 FITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTSSIKALGTVTLPISFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           |||||| ||| |:|| |||||||||||| |||||:|||:  | | |  |||::||:|||
Db     137 GGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDASGYFIASRLIPSINKV 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           :  :|||||  |||||| |||     ||   |||||:|:|||||||||||:||||:| || |||
Db     197 SITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLNDWNFPVSSDSLSYNKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           :| || | |:||||||||| | | | :  |:  | ||||| | || |||||| ||  ||
Db     257 SSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVGSSLQSKPFNLRLRGYN 315

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           ||:| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     316 NSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 375

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 431


RESULT   5
ALS4_CANAL
ID   ALS4_CANAL     STANDARD;      PRT;   469 AA.
AC   O74660;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 4 precursor (Fragment).
GN   ALS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF024584; AAC64239.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   >469       AGGLUTININ-LIKE PROTEIN 4.
FT   NON_TER     469    469
SQ   SEQUENCE   469 AA;  49604 MW;  0EDCAB19B89EFCB1 CRC64;


  Query Match             73.1%;  Score 1600;  DB 1;  Length 469;
  Best Local Similarity   71.2%;  Pred. No. 2e-101;

  Matches  296;  Conservative   42;  Mismatches   78;  Indels    0;  Gaps    0;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || |||||:|||||||:|||:| ::||  ||| ||:|||||| :|: ||||||:||||||
Db      17 AKVITGVFNSFNSLTWANAASYPYRGPATPTWTAVIGWSLDGATASAGDTFTLDMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : | |||:|| |||  ||||   | |||||||:::|||   : :  || |||||| :|:|
Db      77 FITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTADTKAIGTVTLPFSFSV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           ||:||  || :|:|||||  |||||||||  ||  |:|||||  |  :  ||::|||::
Db     137 GGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVASSDRILLSRILPSLSQA 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           ||:  :| |||||||||||::     |||| :|:||: |||||||:||||||||||
Db     197 VNLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           ||  : : :|||||||||:|||||| |: ||: ||| | | |:      ||  | ||
Db     257 TSTSVLVTFQNVPAGYRPFVDAYISATRVSSYTMQYTNIYACVGAASVDDSFTHTWRGYS 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           || |||||| || |||||||||||||||||  ||||||||||||||||||||||||||||||
Db     317 NSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSDTDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432


RESULT   6
SAG1_YEAST
ID   SAG1_YEAST     STANDARD;      PRT;   650 AA.
AC   P20840;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alpha-agglutinin precursor (AG-alpha-1).
GN   SAG1 OR AGA1 OR YJR004C OR J1418.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90014768; PubMed=2677665;
RA   Lipke P.N., Wojciechowicz D., Kurjan J.;
RT   "AG alpha 1 is the structural gene for the Saccharomyces cerevisiae
RT   alpha-agglutinin, a cell surface glycoprotein involved in cell-cell
RT   interactions during mating.";
RL   Mol. Cell. Biol. 9:3155-3165(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=90005993; PubMed=2676603;
RA   Hauser K., Tanner W.;
RT   "Purification of the inducible alpha-agglutinin of S. cerevisiae and
RT   molecular cloning of the gene.";
RL   FEBS Lett. 255:290-294(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RA   de Haan M., Smits P.H.M., Grivell L.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RX   MEDLINE=96064684; PubMed=7592821;
RA   Chen M.-H., Shen Z.-M., Bobin S., Kahn P.C., Lipke P.N.;
RT   "Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for
RT   a yeast cell wall protein with multiple immunoglobulin-like domains
RT   with atypical disulfides.";
RL   J. Biol. Chem. 270:26168-26177(1995).
```

```
CC   -!- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT
CC       TO FACILITATE MATING. SACCHAROMYCES CEREVISIAE A AND ALPHA CELLS
CC       EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ
CC       AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO
CC       PROMOTE CELLULAR AGGREGATION DURING MATING.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POSSIBLE).
CC   -!- INDUCTION: BY EXPOSITION TO PHEROMONE (A-FACTOR) SECRETED BY THE
CC       OPPOSITE MATING TYPE CELLS (TYPE A).
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED.
CC   -!- SIMILARITY: TO C.ALBICANS ALS1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28164; AAA34417.1; -.
DR   EMBL; X16861; CAA34752.1; -.
DR   EMBL; X87611; CAA60926.1; -.
DR   EMBL; Z49504; CAA89526.1; -.
DR   PIR; S22835; S22835.
DR   SGD; S0003764; SAG1.
KW   Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.
FT   SIGNAL        1     19
FT   CHAIN        20    650       ALPHA-AGGLUTININ.
FT   DOMAIN      278    350       ACIDIC REGION, PROBABLY IMPORTANT IN
FT                                BINDING TO AGGLUTININS OF TYPE A CELLS.
FT   DOMAIN      339    423       2 X 40 AA TANDEM REPEATS.
FT   REPEAT      339    378       1.
FT   REPEAT      384    423       2.
FT   DISULFID     97    114
FT   DISULFID    202    300
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    109    109       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    282    282       O-LINKED (MAN. . .).
FT   CARBOHYD    289    289       O-LINKED (MAN. . .).
FT   CARBOHYD    299    299       O-LINKED (MAN. . .).
FT   CARBOHYD    303    303       O-LINKED (MAN. . .).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    307    307       O-LINKED (MAN. . .).
FT   CARBOHYD    308    308       O-LINKED (MAN. . .).
FT   CARBOHYD    311    311       O-LINKED (MAN. . .).
FT   CARBOHYD    314    314       O-LINKED (MAN. . .).
FT   CARBOHYD    315    315       O-LINKED (MAN. . .).
FT   CARBOHYD    316    316       O-LINKED (MAN. . .).
FT   CARBOHYD    329    329       O-LINKED (MAN. . .).
FT   CARBOHYD    331    331       O-LINKED (MAN. . .).
FT   CARBOHYD    334    334       O-LINKED (MAN. . .).
FT   CARBOHYD    335    335       O-LINKED (MAN. . .).
FT   CARBOHYD    338    338       O-LINKED (MAN. . .).
FT   CARBOHYD    339    339       O-LINKED (MAN. . .).
FT   CARBOHYD    340    340       O-LINKED (MAN. . .).
FT   CARBOHYD    341    341       O-LINKED (MAN. . .).
FT   CARBOHYD    342    342       O-LINKED (MAN. . .).
FT   CARBOHYD    345    345       O-LINKED (MAN. . .).
FT   CARBOHYD    346    346       O-LINKED (MAN. . .).
FT   CARBOHYD    349    349       O-LINKED (MAN. . .).
FT   CARBOHYD    350    350       O-LINKED (MAN. . .).
FT   CARBOHYD    364    364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    460    460       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    485    485       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    614    614       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    449    449       S -> P (IN REF. 1).
FT   CONFLICT    556    556       K -> E (IN REF. 1).
FT   CONFLICT    581    581       V -> L (IN REF. 1).
```

```
SQ   SEQUENCE   650 AA;  70339 MW;  8BBF7A1C44C93C2B CRC64;


  Query Match           14.1%;  Score 308.5;  DB 1;  Length 650;
  Best Local Similarity   25.8%;  Pred. No. 9e-14;
  Matches  116;  Conservative   79;  Mismatches  190;  Indels   65;  Gaps   20;

Qy      10 SFNSLTWSN------AANYAFKGPGYPTWNAVLGWSL-DGTSANPGDTFTLNMPCVFK-- 60
           : | :|:||       ||   | |    | |   :|: | :|   || |||:|| |::
Db      21 NINDITFSNLEITPLTAN---KQPD-QGWTATFDFSIADASSIREGDEFTLSMPHVYRIK 76

Qy      61 -YTTSQTSVDLTADGVKYATCQFYSGE-----EFTTFSTLTCTVNDALKSSIKAFGTVTL 114
            :|||:    |||  :   | : | :      | |||   |||  : | |      |::|
Db      77 LLNSSQTATISLADGTEAFKC-YVSQQAAYLYENTTF---TCTAQNDLSSYNTIDGSITF 132

Qy     115 PIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY---LYAS 171
           : |: ||:    :||::| | :|   |   :   |:   | |     ||:|:   :: |
Db     133 SLNFSDGGSSYEYELENAKFFKSGPMLVKLGNQMSDV---VNF-----DPAAFTENVFHS 184

Qy     172 RVMPSLNKVTTLFVAPQCENGYTSG---TMGFSSSNGDVAIDCSNIHIGITKGLNDWNYP 228
                 :  :   | |||  |    : : ||| :| :|||:: :  :   ||| :|
Db     185 GRSTGYGSFESYHLGMYCPNGYFLGGTEKIDYDSSNNNVDLDCSSVQVYSSNDFNDWWFP 244

Qy     229 VSSESFSYTKTCTSNGIQIKY-QNVPAGYRPFIDAYIS-ATDVNQYTLAYTNDYTC---- 282
           |    :   ||  : : |   : :  |   :::|  |   :||    |    |||
Db     245 QSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNALQSLPANVNTIDHALEFQYTCLDTI 304

Qy     283 AGSRLQSKPFTLR----WTGYKNSDAGSNGIVIVATTRTVTD------STTAVTTLPFNP 332
           | :   ::  | |    : |      | :    |  ||  :|      ||  :::|:
Db     305 ANTTYATQFSTTREFIVYQGRNLGTASAKSSFISTTTDLTSINTSAYSTGSISTV---- 360

Qy     333 SVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTT------- 385
               | |::  : ||:   |    |:  : :|:     : : |    |||:
Db     361 ETGNRTTSEVISHVVTTSTKLSPTATTSLTIAQTSIYSTDSNITVGTDIHTTSEVISDVE 420

Qy     386 TVTSEWTGTITTT-TTRTNPTDSIDTVVVQ 414
           |:: |   |:    |: |  | :::| : |
Db     421 TISRETASTVVAAPTSTTGWTGAMNTYISQ 450


RESULT   7
YAG3_YEAST
ID   YAG3_YEAST     STANDARD;      PRT;  1322 AA.
AC   P39712;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Hypothetical 138.1 kDa protein in FLO9-GDH3 intergenic precursor.
GN   YAL063C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC   -!- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U12980; AAC04971.1; -.
DR   SGD; S0000059; YAL063C.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 13.
KW   Hypothetical protein; Glycoprotein; Transmembrane; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25   1322       HYPOTHETICAL PROTEIN YAL063C.
FT   TRANSMEM   366    388       POTENTIAL.
FT   TRANSMEM   754    775       POTENTIAL.
FT   CARBOHYD   135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   257    257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   270    270       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   464    464       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   509    509       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   644    644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   734    734       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   888    888       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1322 AA;  138072 MW;  AADFD1FF13267CEA CRC64;

  Query Match             8.1%;  Score 177;  DB 1;  Length 1322;
  Best Local Similarity   23.3%;  Pred. No. 0.00017;
  Matches  112;  Conservative   53;  Mismatches  181;  Indels  134;  Gaps   26;

Qy      21 NYAFKGPGYPTWNAVLG-WSLD--GTSANPGDTFTLNM----------PCVFKYTTSQTS 67
           |:  || |  : | ::  || |  |    | :  || |            ||: |   |
Db     104 NWGCKGIGACSNNPIIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDS 162

Qy      68 VDLTADG-VKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAF---------GTV----- 112
             |:  | : :  |     :|    ::   |:|      || :          |||
Db     163 AILSVGGSIAFECC----AQEQPPITSTNFTIN-----GIKPWNGSPPDNITGTVYMYAG 213

Qy     113 ----------------TLPIAFNV-GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDV 155
                           ||||:  :  ||  | | |           ||   | ::|
Db     214 FYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEG--------YVYTF---DNNLS--- 259

Qy     156 EFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIH 215
           :  :  ||| |   |  : :    |  | :  |    :|| |  :    :|
Db     260 QPNCTIPDPSNYT-VSTTITTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTA 318

Qy     216 IGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLA 275
             |      ||   :| |   |    :||::     :       |    :||     |
Db     319 STIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEP 372

Qy     276 YTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFN 331
           : : :|   : |              :  |:||:     :   || | :|||:||  |:|
Db     373 WNSTFTSTSTEL------------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWN 419

Qy     332 PSVDKTKT----------------IEILQPIPTTTITT-------SYVGVTTSYLTKTA 367
             :   | |               | |  |   ||  |        ::   :|   | |
Db     420 DTFTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTG 479

Qy     368 ----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVV 413
               | ||  :::  |  ||  | |  |  | |      || | ||  |||  :|::|
Db     480 TNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTD--ETIIV 536


RESULT   8
FLO1_YEAST
ID   FLO1_YEAST     STANDARD;      PRT;  1537 AA.
AC   P32768;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Flocculation protein FLO1 precursor (Flocculin 1).
GN   FLO1 OR YAR050W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94262325; PubMed=8203162;
RA   Watari J., Takata Y., Ogawa M., Sahara H., Koshino S., Onnela M.-L.,
RA   Airaksinen U., Jaatinen R., Penttilae M., Keraenen S.;
RT   "Molecular cloning and analysis of the yeast flocculation gene FLO1.";
RL   Yeast 10:211-225(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN   [3]
RP   PRELIMINARY SEQUENCE FROM N.A.
RX   MEDLINE=93289821; PubMed=8511970;
RA   Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,
RA   Steensma H.Y.;
RT   "Sequence of the open reading frame of the FLO1 gene from
RT   Saccharomyces cerevisiae.";
RL   Yeast 9:423-427(1993).
RN   [4]
RP   REVIEW.
RX   MEDLINE=96076625; PubMed=7502576;
RA   Teunissen A.W., Steensma H.Y.;
RT   "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT   constitute a new subtelomeric gene family.";
RL   Yeast 11:1001-1013(1995).
CC   -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78160; CAA55024.1; -.
DR   EMBL; L28920; AAC09499.1; ALT_SEQ.
DR   PIR; S31230; S31230.
DR   SGD; S0000084; FLO1.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 18.
KW   Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25      ?       FLOCCULATION PROTEIN FLO1.
FT   PROPEP       ?   1537       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN     274   1240       SER/THR-RICH.
FT   DOMAIN     278   1087       18 X 45 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     278    322       1-1.
FT   REPEAT     323    367       1-2.
FT   REPEAT     368    412       1-3.
FT   REPEAT     413    457       1-4.
FT   REPEAT     458    502       1-5.
FT   REPEAT     503    547       1-6.
FT   REPEAT     548    592       1-7.
FT   REPEAT     593    637       1-8.
FT   REPEAT     638    682       1-9.
FT   REPEAT     683    727       1-10.
```

```
FT   REPEAT      728    772       1-11.
FT   REPEAT      773    817       1-12.
FT   REPEAT      818    862       1-13.
FT   REPEAT      863    907       1-14.
FT   REPEAT      908    952       1-15.
FT   REPEAT      953    997       1-16.
FT   REPEAT      998   1042       1-17.
FT   REPEAT     1043   1087       1-18.
FT   DOMAIN     1118   1157       2 X 20 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1118   1137       2-1.
FT   REPEAT     1138   1157       2-2.
FT   DOMAIN     1226   1392       3 X 51 AA APPROXIMATE REPEATS.
FT   REPEAT     1226   1276       3-1.
FT   REPEAT     1291   1341       3-2.
FT   REPEAT     1342   1392       3-3.
FT   DOMAIN     1408   1434       3 X 9 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1408   1416       4-1.
FT   REPEAT     1417   1425       4-2.
FT   REPEAT     1426   1434       4-3.
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    374    374       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    509    509       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    644    644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    734    734       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1537 AA;  160722 MW;  992650C6BE9A8CEA CRC64;
```

```
  Query Match             8.1%;  Score 177;  DB 1;  Length 1537;
  Best Local Similarity   23.0%;  Pred. No. 0.00021;
  Matches  112;  Conservative   47;  Mismatches  181;  Indels  146;  Gaps   25;

Qy      21 NYAFKGPG-YPTWNAVLGWSLD--GTSANPGDTFTLNM----------PCVFKYTTSQ-- 65
           |:  || |       :  || |  |    | :  || |                ||: |
Db     104 NWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDS 162

Qy      66 ---------------------TSVDLTADGVK----------YATCQFYSGEEFTTFSTL 94
                               || : | ||:|           |   |:|  :
Db     163 AILSVGGATAFNCCAQQQPPITSTNFTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---- 218

Qy      95 TCTVNDALKSSIKAFGT----VTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKD 150
                 : |:  ::||    ||||      ||  | | |           :|:|
Db     219 -----KVVYSNAVSWGTLPISVTLP-----DGTTVSDDFEG--------YVYSFDD---- 256

Qy     151 ISIDVEFEKSTV-DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAI 209
              |:     || ||| |   |    :    |  | :   |    :|| |  :    : |
Db     257 ---DLSQSNCTVPDPSNYA-VSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVI 312

Qy     210 DCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDV 269
                 |      ||   :| | |   |    :||::    :       |    :||
Db     313 RTPTTASTIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETII------VIRTPTTATTA 366

Qy     270 NQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVT 326
              |  : : :|   : |              :  |:||:     :   || | :|||:|
Db     367 ITTTEPWNSTFTSTSTEL------------TTVTGTNGLPTDETIIVIRTPTTATTAMT 413

Qy     327 TL-PFNPSVDKTKT----------------IEILQPIPTTTITT-------SYVGVTTS 361
           |  |:| :   | |                | |  |   ||  |       ::   :|
Db     414 TTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTE 473

Qy     362 YLTKTA----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTTTRTN--PTDS 407
             | |      |  ||  :::  |   ||  | |  |  | |      || | ||  |||
Db     474 MTTVTGTNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTD- 531

Qy     408 IDTVVV 413
               :|::|
```

```
Db     532 -ETIIV 536
```

```
RESULT   9
YQU3_CAEEL
ID   YQU3_CAEEL     STANDARD;      PRT;  1251 AA.
AC   Q09550;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical 133.5 kDa protein F26C11.3 in chromosome II.
GN   F26C11.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Matthews P.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47072; CAA87369.1; -.
DR   WormPep; F26C11.3; CE01561.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 1.
DR   SMART; SM00032; CCP; 1.
KW   Hypothetical protein.
FT   DOMAIN      120    414       SER/THR-RICH.
SQ   SEQUENCE   1251 AA;  133498 MW;  2B959ECA03B9954A CRC64;
```

```
  Query Match             7.8%;  Score 171;  DB 1;  Length 1251;
  Best Local Similarity   23.8%;  Pred. No. 0.00041;
  Matches  107;  Conservative   60;  Mismatches  154;  Indels  128;  Gaps   23;

Qy      27 PGYPTWNAVLGWSLDGTS---------------ANPGDTFTLNMPCVFKYTTSQTSVDLT 71
           ||  |:|    |   ||:                | |  | |: |  ::  :|::|  : |
Db     609 PGTTTYN----WPTGGTTRMLPSGEIILSESLIAYPNCT-TVLMQLIYTPSTNKTRTETT 663

Qy      72 AD--GVKYATCQFYSGEEFT-----TFSTLTCTVN--------------DALKSSIKAF- 109
           |  | | :   |  :|:     | |: | | |                |  |: ||
Db     664 TDTEGCKKTSTISSSSSKFSITPTPTPSSGTTTYNWPTGGTTRTLPSGEIILSESLIAFQ 723

Qy     110 --GTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 167
             || : : :|      ||:          | | |  :| |  |   :   :   |:
Db     724 NCTTVLMQLIYN-----PSTN-------KTRTETTTDAEGCKKTSSTSKISTTPTSPT-- 769

Qy     168 LYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNY 227
             :|:  |:   :|| :  |    | |: |:     :|::
Db     770 --SSKPTPTSTSMTTTYNWP---TGGTTRTL----PSGEI-------------------- 800

Qy     228 PVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVN--QYTLAYTNDYTCAGS 285
           : |||    | ||:  :|: |   |:  :   :   : ||    : | : :   |   |
Db     801 -ILSESLIAYKNCTTVLMQLIYN--PSTNKTRTE---TTTDAQGCKATSSTSLKPTSPSS 854

Qy     286 RLQSKP-FTLRW-TGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKT---- 339
           | |  |  | ||       |  |::  :     : || :  | :|||  :||:|
Db     855 STASPPTTTYNWPTGGTTRTLPSGEIILSESLIAYKNCTTVLMQLIYNPSTNKTRTETTT 914

Qy     340 ------IEILQPIPTTTITTSYVGVTTSYLTKTAPIGE-----------TATVIVDVPY 381
                 |  | |   ||||:|    |   |:| | ||              ||:: : |
Db     915 DAQGCKATITTPTP---ITTTY-NWPTGGTTRTLPSGEIILSESLIAYKNCTTVLMQLIY 970

Qy     382 H-----TTTTVTSEWTGTITTTTTRTNPT 405
```

```
          :     | |  ||:  |   |:||:|  |
Db    971 NPSTNKTRTETTSDAQGCKATSTTQTPTT 999


RESULT 10
MUC1_XENLA
ID   MUC1_XENLA     STANDARD;      PRT;   662 AA.
AC   Q05049;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   Integumentary mucin C.1 (FIM-C.1) (Fragment).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=93077556; PubMed=1447205;
RA   Hauser F., Hoffmann W.;
RT   "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT   C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT   polymorphism.";
RL   J. Biol. Chem. 267:24620-24624(1992).
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 6 P-TYPE (TREFOIL) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L02115; AAA74725.1; -.
DR   PIR; A45155; A45155.
DR   HSSP; P01359; 2PSP.
DR   InterPro; IPR000519; P_trefoil.
DR   Pfam; PF00088; trefoil; 6.
DR   SMART; SM00018; P; 6.
DR   PROSITE; PS00025; P_TREFOIL; 6.
KW   Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN       81    144       8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT                                ALA/THR-RICH.
FT   REPEAT       81     88       1-1.
FT   REPEAT       89     96       1-2.
FT   REPEAT       97    104       1-3.
FT   REPEAT      105    112       1-4.
FT   REPEAT      113    120       1-5.
FT   REPEAT      121    128       1-6.
FT   REPEAT      129    136       1-7.
FT   REPEAT      137    144       1-8.
FT   DOMAIN      161    202       P-TYPE 1.
FT   DOMAIN      218    301       8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT   REPEAT      218    224       2-1.
FT   REPEAT      225    239       2-2.
FT   REPEAT      240    249       2-3.
FT   REPEAT      250    259       2-4.
FT   REPEAT      260    275       2-5.
FT   REPEAT      276    287       2-6.
FT   REPEAT      288    294       2-7.
FT   REPEAT      295    301       2-8.
FT   DOMAIN      306    347       P-TYPE 2.
FT   DOMAIN      353    394       P-TYPE 3.
FT   DOMAIN      402    522       12 X APPROXIMATE TANDEM REPEATS,
FT                                THR-RICH.
FT   REPEAT      402    411       3-1.
FT   REPEAT      412    419       3-2.
FT   REPEAT      420    431       3-3.
FT   REPEAT      432    443       3-4.
FT   REPEAT      444    453       3-5.
FT   REPEAT      454    460       3-6.
FT   REPEAT      461    472       3-7.
FT   REPEAT      473    479       3-8.
FT   REPEAT      480    491       3-9.
FT   REPEAT      492    498       3-10.
FT   REPEAT      499    515       3-11.
FT   REPEAT      516    522       3-12.
FT   DOMAIN      525    566       P-TYPE 4.
FT   DOMAIN      572    613       P-TYPE 5.
FT   DOMAIN      620    661       P-TYPE 6.
FT   DISULFID    162    188       BY SIMILARITY.
FT   DISULFID    172    187       BY SIMILARITY.
FT   DISULFID    182    199       BY SIMILARITY.
FT   DISULFID    307    333       BY SIMILARITY.
FT   DISULFID    317    332       BY SIMILARITY.
FT   DISULFID    327    344       BY SIMILARITY.
FT   DISULFID    354    380       BY SIMILARITY.
FT   DISULFID    364    379       BY SIMILARITY.
FT   DISULFID    374    391       BY SIMILARITY.
FT   DISULFID    526    552       BY SIMILARITY.
FT   DISULFID    536    551       BY SIMILARITY.
FT   DISULFID    546    563       BY SIMILARITY.
FT   DISULFID    573    599       BY SIMILARITY.
FT   DISULFID    583    598       BY SIMILARITY.
FT   DISULFID    593    610       BY SIMILARITY.
FT   DISULFID    621    647       BY SIMILARITY.
FT   DISULFID    631    646       BY SIMILARITY.
FT   DISULFID    641    658       BY SIMILARITY.
FT   VARIANT     276    276       K -> E.
FT   VARIANT     354    354       C -> R.
FT   VARIANT     415    415       T -> A.
SQ   SEQUENCE   662 AA;  67774 MW;  F085277F1ED2FD40 CRC64;


  Query Match             7.6%;  Score 167.5;  DB 1;  Length 662;
  Best Local Similarity   24.3%;  Pred. No. 0.00033;
  Matches   94;  Conservative   28;  Mismatches  166;  Indels   99;  Gaps   16;

Qy      42 GTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDA 101
           | |     | |   |   :  |: |:   |   |                 |  :| |
Db     206 GHSHEEHTTTTTKAPTTIQIATTTTTPTTTTTTTK----------ATPTTTTTTKATPT 254

Qy     102 LKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKST 161
           :: ||  | | |          :|    :|  | | |                    :|
Db     255 TTTTTKATTTTTTP----------TTTTTTTKATTTPTTT-----------------TT 286

Qy     162 VDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKG 221
             |:          :  | ||           |||      |  :   ||   : |||:
Db     287 TTPTT---------TTTKATTTTTT-------TSGECKMEPSKRE---DCG--YSGITES 325

Qy     222 LNDWNYPVSSESFSYTKTCTSNGIQI------KYQNVPAGYRPFIDAYISATDVNQYTLA 275
                   |   || |     |:       | |  |:|      |:|    |
Db     326 QCRTKGCCFDSSIPQTKWCFYTLSQVADCKVEPSQRVDCGFRG-----ITADQCRQKNCC 380

Qy     276 YTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVD 335
           :  | : :|:          :|  |  |   ,     || |:| |  ||  || |  :
Db     381 F--DSSISGT---------KWCFYSTSQ--------VAATKTTTPTT--TTTPTTTTTT 419

Qy     336 K----TKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEW 391
           |    | |      |  ||| ||:    ||:  | | |   |       |  |||| |:
Db     420 KATTTTPTTTTTTPTTTTTTTTT-TKATTTTPTTTTPTTTTTKATTTTPTTTTTTPTTTT 478

Qy     392 TGTITT--TTTRTNPTDSIDTVVVQVP 416
           |   ||  ||| | || :        |
Db     479 TKATTTTPTTTTTTPTTTTTKATTTTP 505
```

```
RESULT  11
FLO5_YEAST
ID   FLO5_YEAST     STANDARD;      PRT;  1075 AA.
AC   P38894;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Flocculation protein FLO5 precursor (Flocculin 5).
GN   FLO5 OR YHR211W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [2]
RP   REVIEW.
RX   MEDLINE=96076625; PubMed=7502576;
RA   Teunissen A.W., Steensma H.Y.;
RT   "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT   constitute a new subtelomeric gene family.";
RL   Yeast 11:1001-1013(1995).
CC   -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00029; AAB69731.1; -.
DR   PIR; S48992; S48992.
DR   SGD; S0001254; FLO5.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 8.
KW   Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       FLOCCULATION PROTEIN FLO5.
FT   PROPEP        ?   1075       REMOVED IN MATURE FORM (POTENTIAL).
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    749    749       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1075 AA;  111981 MW;  D151B370B60C8D9F CRC64;


  Query Match             7.4%;  Score 161.5;  DB 1;  Length 1075;
  Best Local Similarity   22.2%;  Pred. No. 0.0015;
  Matches  110;  Conservative   57;  Mismatches  209;  Indels  119;  Gaps   22;

Qy      21 NYAFKGPG-YPTWNAVLGWSLD--GTSANPGDTFTLNM------PCVFKYTTSQTSVDLT 71
           |:  || |       :  || |  |    | :  || |       |    || |  :|| :
Db     104 NWGCKGMGACSNSQGIAYWSTDLFGFYTTPTNV-TLEMTGYFLPPQTGSYTFSFATVDDS 162

Qy      72 A-----DGVKYATCQFYSGEEFTTFSTLTCTVN---------------------------- 99
           |       : :  |     :|     ::    |:|
Db     163 AILSVGGSIAFECC----AQEQPPITSTNFTINGIKPWDGSLPDNITGTVYMYAGYYYPL 218

Qy     100 DALKSSIKAFGTVTLPIAFNVGGTGSST----------DLEDSKC-------FTAGTNTV 142
             : |:  ::||: : :    | | |             ||  | |         |  | |
Db     219 KVVYSNAVSWGTLPISVELPDGTTVSDNFEGYVYSFDDDLSQSNCTIPDPSIHTTSTITT 278

Qy     143 -------TFNDGDKDISIDVEFEKSTVDPSAYLY-----ASRVMPSLNKVTTLFVAPQCE 190
                 ||      :::   :      | :  :      || :  :    |  | :   |
Db     279 TTEPWTGTFTSTSTEMTTITDTNGQLTDETVIVIRTPTTASTITTTTEPWTGTFTSTSTE 338

Qy     191 NGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQ 250
               :|| |  :   :  :|          ||      |   :| |   |     :|| |   :
Db     339 MTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNG-QPTDE 397

Qy     251 NVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSD------- 303
           |      |  :  |: |      |  :|  :|    : :     |:  |  : :|
Db     398 TVIVIRTPTSEGLITTT-----TEPWTGTFTSTSTEVT----TITGTNGQPTDETVIVIR 448

Qy     304 -AGSNGIVIVAT---TRTVTDSTTAVTTLPFNPSVDKTKTIEILQP------IPTT---- 349
              | |::   |    | | | ::| :||:         :|: :::       | ||
Db     449 TPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPW 508

Qy     350 --TITTSYVGVTTSYLTKTAPIGETATVI-VDVPYHTTTTVTSEWTGTIT------TTTT 400
             | |::   |||   |   |  ||  ||         || |  ||||  |      || |
Db     509 TGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVT 568

Qy     401 RTN--PTDSIDTVVV 413
             ||  |||  :||:|
Db     569 GTNGQPTD--ETVIV 581


RESULT  12
CHI3_CANAL
ID   CHI3_CANAL     STANDARD;      PRT;   567 AA.
AC   P40954;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Chitinase 3 precursor (EC 3.2.1.14).
GN   CHT3.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10261;
RX   MEDLINE=95223977; PubMed=7708682;
RA   McCreath K.J., Specht C.A., Robbins P.W.;
RT   "Molecular cloning and characterization of chitinase genes from
RT   Candida albicans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:2544-2548(1995).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC       acetyl-D-glucosamine polymers of chitin.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   ----------;---------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U15801; AAA68016.1; -.
DR   HSSP; P23472; 2HVM.
```

```
DR   InterPro; IPR001579; Chitinase_2.
DR   Pfam; PF00192; chitinase_2; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;
KW   Chitin-binding; Multigene family.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    567       CHITINASE 3.
FT   ACT_SITE    157    157       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DOMAIN      319    436       SER/THR-RICH.
SQ   SEQUENCE   567 AA;  60060 MW;  DD843126F65E22C2 CRC64;


  Query Match             7.3%;  Score 160;  DB 1;  Length 567;
  Best Local Similarity   20.3%;  Pred. No. 0.00087;
  Matches   94;  Conservative   46;  Mismatches  138;  Indels  184;  Gaps   17;

Qy      10 SFNSLTW----------------------SNAANYAFKGP-----------------GY 29
           |||  ||                        | ||   :  |                 |
Db     225 SFNWDTWLNYAETDSPNKNIKLFVGVPASSRAAGSGYNDPSAVSQYLTSDILNSKYFGGI 284

Qy      30 PTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFT 89
           |:  |||    : |    |  ||  : |  :                        ||| |
Db     285 SMWDVSAGWS----NTNSNGNFVENMKAIVKKAS--------------------PGEETT 320

Qy      90 TFSTLT------CTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVT 143
           : || |       |:: :  ||  :  : |   : ::  | |||    |   |: : :
Db     321 SSSTTTTTTTTSTTISSSSSSSKTSKTSTTSTTSSSISSTTSSTTSSTSSSSTSSSTS-- 378

Qy     144 FNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSS 203
                       ||   |          : ::::|   ||      :| |:  |:|
Db     379 ---------------STTSSST---------TSSQISTTSTAPTSSTSLSSSTISTSAS 413

Qy     204 NGDVAIDCSNIHIGITKGLNDWNYPVSSESFS-YTKTCTSNGIQIKYQNVPAGYRPFIDA 262
           |           :|          |||:    | :  |: | |   :   |
Db     414 TSDTT--------SVT----------SSETTPVVTPSSLSSAITIPGDSTTTG------- 448

Qy     263 YISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDST 322
           || :                :|| |                    :||  :: ||
Db     449 -ISKSS------------------STKPAT------------------STTSALSSST 469

Qy     323 TAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYH 382
           | | |:| :  :  | |       |   || |     |  || :       |
Db     470 TTVATIPDDKEIINTPTDTETTSKPPAIITESDATTITQNLTPSTTTKNVKT-------- 521

Qy     383 TTTTVTSEWT------GTITTT----TTRTNPTDSIDTVVVQ 414
           |:| : :||        |:|||    ||||:    |:||  :
Db     522 TSTNIVTEWVWAPTTLRTLTTTYQILTTRTH----IETVFAE 559


RESULT  13
GUX1_ASPAC
ID   GUX1_ASPAC     STANDARD;      PRT;   540 AA.
AC   O59843;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-
DE   beta-cellobiohydrolase I) (Beta-glucancellobiohydrolase I).
GN   CBHI.
OS   Aspergillus aculeatus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5053;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NO. F-50;
RA   Takada G., Kawaguchi T., Sumitani J., Arai M.;
RT   "Cloning, nucleotide sequence, and transcriptional analysis of
RT   Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhI) gene.";
RL   J. Ferment. Bioeng. 85:1-9(1998).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC           FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC           SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB002821; BAA25183.1; -.
DR   HSSP; P00725; 1AZ6.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; Glyco_hydro_7.
DR   Pfam; PF00734; CBD_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   PRINTS; PR00734; GLHYDRLASE7.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; Glyco_hydro_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; FALSE_NEG.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    540       EXOGLUCANASE I.
FT   DOMAIN       23    459       CATALYTIC.
FT   DOMAIN      460    505       LINKER.
FT   DOMAIN      506    540       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    234    234       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    239    239       PROTON DONOR (BY SIMILARITY).
FT   DISULFID    513    529       BY SIMILARITY.
FT   DISULFID    524    539       BY SIMILARITY.
FT   CARBOHYD    434    434       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    458    458       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   540 AA;  57099 MW;  002D7FD28DF194D0 CRC64;


  Query Match             7.1%;  Score 156;  DB 1;  Length 540;
  Best Local Similarity   21.4%;  Pred. No. 0.0015;
  Matches  112;  Conservative   57;  Mismatches  163;  Indels  192;  Gaps   26;

Qy       4 ITGVFDSFNSLTWSN---------AANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLN 54
           : |  : ::  || :           |:  | :|  | :  |             | :  ||
Db      66 VGGYTNCYSGNTWDSSICSTDTTCASECALEGATYESTYGV---------TTSGSSLRLN 116

Qy      55 MPCVFKYTTSQTSVD----LTADGVKYATCQFYSGEEFT---TFSTLTCTVNDALK-SSI 106
               | || ::     | ||   | | : ::  |||     | | | :| ||   |:
Db     117 ----FVTTASQKNIGSRLYLLADDSTYETFKLFN-REFTFDVDVSNLPCGLNGALYFVSM 171

Qy     107 KAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSA 166
           | | |:     |  |    |   ||:|      : | ||  :|     :| |: | :|
Db     172 DADGGVS-RFPTNKAGAKYGTGYCDSQC---PRDLKFIDGQANIE---GWEPSSTDVNA 223

Qy     167 YL-YASRVMPSL-----NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITK 220
                 | :     | ::: | |  |::     ||      ||    |   :
Db     224 GTGNHGSCCPEMDIWEANSISSAFTAHPCDS--VQQTM----CTGDT---CGGTY----- 269

Qy     221 GLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDY 280
                   |  :  |: ||  :|                          | | |    || |
Db     270 ---------SDTTDRYSGTCDPDG---------------------CDFNPYRFGNTNFY 298

Qy     281 TCAGSRLQSKPFTL--RWTGYKNSDAGS----------NGIVI--------VATTRTVT- 319
                 :  |||||:  ::  :  :| |:          ||:||         |:  ::|
Db     299 GPGKTVDNSKPFTVVTQFITHDGTDTGTLTEIRRLYVQNGVVIGNGPSTYTAASGNSITE 358
```

```
Qy    320 ------------------------------------------------DSTTAVTTLP 329
                                                           ||    |:
Db    359 SFCKAEKTLFGDTNVFETHGGLSAMGDALGDGMVLVLSLWDDHAADMLWLDSDYPTTSCA 418

Qy    330 FNPSVDKTKTIEILQPIPTTTITTSYV--GVTTSYLT----KTAPIGET---------- 372
          :| | :      ||||  :||    ||:|   |  :  |
Db    419 SSPGVAR-------GTCPTTTGNATYVEANYPNSYVTYSNIKFGTLNSTYSGTSSGGSSS 471

Qy    373 -ATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQV 415
           :| :      ||   ||  :   ||||::|: | |  | | |:
Db    472 SSTTL-----TTKASTSTTSSKTTTTTSKTSTTSSSSTNVAQL 509
```

```
RESULT  14
AMYH_YEAST
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN     210   1367       SER/THR-RICH.
FT   CARBOHYD   817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;
```

```
  Query Match             7.1%;  Score 154.5;  DB 1;  Length 1367;
  Best Local Similarity   23.6%;  Pred. No. 0.006;
  Matches  115;  Conservative   40;  Mismatches  184;  Indels  149;  Gaps   22;

Qy     15 TWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQ--TSVDLTA 72
          | |:::|     |      : : :|    | : | | |    |   ||  ||  |||  |
Db    868 TPSSSSNITSSAP------SSIPFSSTTESFSTGTTVT---PSSSKYPGSQTETSVSSTT 918

Qy     73 DG--VKYATCQFYSGEEFTTFSTLTCT--VNDALKS----SIKAFGTVTLPIAFNVGGTG 124
          :   |   |    :    || :|  |:   | | ::    | |   | |:|       |
Db    919 ETTIVPTKTTTSVTTPSTTTITTTVCSTGTNSAGETTSGCSPKTV-TTTVPTTTTTSVTT 977

Qy    125 SSTDLEDSKCFTAGTNT---VTFNDGDKDIS---------------------------- 152
          |||    :   : |||:    |     | |:
Db    978 SSTTTITTTVCSTGTNSAGETTSGCSPKTITTTVPCSTSPSETASESTTTSPTTPVTTVV 1037

Qy    153 ----IDVEFEKSTVD-------------PSAYLYASRVMPSLNKVT----TLFVAPQCEN 191
              : |:  ||                |: ||     ||:  ||    |     |
Db   1038 STTVVTTEYSTSTKPGGEITTTFVTKNIPTTYLTTIAPTPSVTTVTNFTPTTITTTVCST 1097

Qy    192 GYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQN 251
          |         ::| |:    ||   :  |         | |: :  ||    |:
Db   1098 G--------TNSAGETTSGCSPKTVTTT-------VPCSTGTGEYTTEATT--------- 1133

Qy    252 VPAGYRPFIDAYISATDVNQYTLAYTND--YTCAGSRLQSKPFTLRWTGYKNSDAGSNGI 309
                :   :: | |   :   ||    |  |   :| |     | |  : | |  :
Db   1134 -------LVTTAVTTTVVTTESSTGTNSAGKTTTGYTTKSVP-----TTYVTTLAPSAPV 1181

Qy    310 V----IVATTRTVTDSTTAVTTLPFNPSVDKTKTI---------------EILQPIPTT 349
                | || | |: : |          ||   |||                :   ||||
Db   1182 TPATNAVPTTITTTECSAATNAAGETTSVCSAKTIVSSASAGENTAPSATTPVTTAIPTT 1241

Qy    350 TITT-SYVG------VTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRT 402
           ||| | ||       || | ||: |           | ||    |      |  | |
Db   1242 VITTESSVGTNSAGETTTGYTTKSIP----------TTYITTLIPGSNGAKNYETVATAT 1291

Qy    403 NPTDSIDT 410
          ||  || |
Db   1292 NPI-SIKT 1298
```

```
RESULT  15
AGA1_YEAST
ID   AGA1_YEAST     STANDARD;      PRT;   725 AA.
AC   P32323;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   A-agglutinin attachment subunit precursor.
GN   AGA1 OR YNR044W OR N3431.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91304412; PubMed=2072914;
RA   Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT   "The AGA1 product is involved in cell surface attachment of the
RT   Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RL   Mol. Cell. Biol. 11:4196-4206(1991).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pohl T.M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC       SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC       CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC       RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC       AGGREGATION DURING MATING.
CC   -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC       A CORE SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60590; AAA34382.1; -.
DR   EMBL; Z71659; CAA96325.1; -.
DR   PIR; S17031; S17031.
DR   PIR; A41258; A41258.
DR   SGD; S0005327; AGA1.
KW   Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
KW   Pheromone response.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    725       A-AGGLUTININ ATTACHMENT SUBUNIT.
FT   DOMAIN      53    493       2 X APPROXIMATE REPEATS.
FT   REPEAT      53    149       1-1.
FT   REPEAT     395    493       1-2.
FT   DOMAIN     182    307       18 X APPROXIMATE TANDEM REPEATS, SER/THR-
FT                               RICH.
FT   REPEAT     182    188       2-1.
FT   REPEAT     189    195       2-2.
FT   REPEAT     196    202       2-3.
FT   REPEAT     203    209       2-4.
FT   REPEAT     210    216       2-5.
FT   REPEAT     217    223       2-6.
FT   REPEAT     224    230       2-7.
FT   REPEAT     231    237       2-8.
FT   REPEAT     238    244       2-9.
FT   REPEAT     245    251       2-10.
FT   REPEAT     252    258       2-11.
FT   REPEAT     259    265       2-12.
FT   REPEAT     266    272       2-13.
FT   REPEAT     273    279       2-14.
FT   REPEAT     280    286       2-15.
FT   REPEAT     287    293       2-16.
FT   REPEAT     294    300       2-17.
FT   REPEAT     301    307       2-18.
SQ   SEQUENCE   725 AA;  73353 MW;  70420C853B0B01F8 CRC64;


  Query Match             7.0%;  Score 153.5;  DB 1;  Length 725;
  Best Local Similarity   21.9%;  Pred. No. 0.0032;
  Matches   95;  Conservative   64;  Mismatches  209;  Indels   65;  Gaps   15;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           |  |: |  : :| | ||           ||       ||  || :|  | |   |
Db     163 ASIISPVTSTLSSTTSSN----------PTTT-----SLSSTSTSPSSTST--SPSSTS 204

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL-----KSSIKAFGTVTLP 115
            ::| ||   ::     ::    |    :| |:|| | : :       :|  :  | | |
Db     205 TSSSSTSTSSSSTSTSSSSTS-TSPSSTSTSSSLTSTSSSSTSTSQSSTSTSSSSTSTSP 263

Qy     116 IAFNVGGTGSSTDLEDSKCFTAGTNTVTF----------------NDGDKDISIDVEFEK 159
           : :   : :||         : |:| ::                :      ||   |
Db     264 SSTSTSSSSTSTSPSSKSTSASSTSTSSYSTSTSPSLTSSSPTLASTSPSSTSISSTFTD 323

Qy     160 STVDPSAYLYASRVMPSL-NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGI 218
           ||    : : :|    || :  | ::  |   :  : :| :|| :  :   :    |
Db     324 STSSLGSSIASSSTSVSLYSPSTPVYSVPSTSSNVATPSM--TSSTVETTVSSQSSSEYI 381

Qy     219 TKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTN 278
           ||       |  | |  :|   | :|:   |      : |:     ::|  : :    |:
Db     382 TKSSISTTIPSFSMSTYFT---TVSGVTTMYTT----WCPYSSESETSTLTSMHETVTTD 434

Qy     279 DYTCA-GSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKT 337
              |   | : |:  :|  :  | |   :  :    :| ||  |    |    : |
Db     435 ATVCTHESCMPSQTTSLITSSIKMS---TKNVATSVSTSTVESSYACSTCAETSHSYSSV 491

Qy     338 KTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITT 397
           :|       || | |:|   :|  |      : ||      ||    |||  | ||:|
Db     492 QTASSSSVTQQTTSTKSWV---SSMTTSDEDFNKHAT----GKYH----VTSSGTSTIST 540

Qy     398 TTTRTNPTDSIDT 410
           : :    | |||:
Db     541 SVSEATSTSSIDS 553


Search completed: October  3, 2002, 16:34:17
Job time: 250 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:29:47 ; Search time 31.78 Seconds
(without alignments)
2264.503 Million cell updates/sec

Title: US-09-715-876-8_COPY_17_432
Perfect score: 2190
Sequence: 1 AKTITGVFDSFNSLTWSNAA.....TTTTRTNPTDSIDTVVVQVP 416

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1922	87.8	1270	3	Q9Y743	Q9y743 candida alb
2	1889	86.3	1047	3	O93865	O93865 candida alb
3	1764.5	80.6	468	3	Q9URQ0	Q9urq0 candida alb
4	1605	73.3	1523	3	Q9HFX4	Q9hfx4 candida alb
5	1604	73.2	469	3	Q9URP8	Q9urp8 candida alb
6	1602.5	73.2	468	3	Q9C471	Q9c471 candida alb
7	1549	70.7	1443	3	Q9Y8F2	Q9y8f2 candida alb
8	1471	67.2	336	3	Q9HF70	Q9hf70 candida dub
9	1191	54.4	338	3	Q9HF72	Q9hf72 candida dub
10	1176	53.7	331	3	Q9HF69	Q9hf69 candida dub
11	997.5	45.5	2297	3	Q9HGK6	Q9hgk6 candida alb
12	990.5	45.2	433	3	Q9Y8F1	Q9y8f1 candida alb
13	981	44.8	353	3	Q9HF71	Q9hf71 candida tro
14	359.5	16.4	126	3	Q9HF49	Q9hf49 candida tro
15	341	15.6	127	3	Q9HF50	Q9hf50 candida tro
16	219	10.0	52	3	O93861	O93861 candida alb
17	217	9.9	1195	3	Q96WU8	Q96wu8 schizosacch
18	199	9.1	1752	2	Q9AE52	Q9ae52 ruminococcu
19	198.5	9.1	948	3	O74346	O74346 schizosacch
20	180	8.2	1498	2	Q9L448	Q9l448 arthrobacte
21	179.5	8.2	956	5	O00908	O00908 cryptospori
22	175.5	8.0	267	3	Q9P6S0	Q9p6s0 schizosacch
23	175.5	8.0	1832	5	O96503	O96503 cryptospori
24	175	8.0	565	3	Q9HDY9	Q9hdy9 schizosacch
25	172	7.9	1217	4	Q9UKW9	Q9ukw9 homo sapien
26	171.5	7.8	456	3	O42840	O42840 schizosacch
27	170	7.8	957	4	O14651	O14651 homo sapien
28	167	7.6	1220	3	Q9C0Y2	Q9c0y2 schizosacch
29	165	7.5	513	4	O43418	O43418 homo sapien
30	163	7.4	648	5	Q95QX0	Q95qx0 caenorhabdi
31	162.5	7.4	567	17	Q96YT2	Q96yt2 sulfolobus
32	156.5	7.1	770	5	Q20908	Q20908 caenorhabdi
33	156	7.1	851	5	O17893	O17893 caenorhabdi
34	155.5	7.1	1893	5	Q9NKC9	Q9nkc9 drosophila
35	155	7.1	597	16	Q9RY11	Q9ry11 deinococcus
36	155	7.1	1180	16	Q9CHH4	Q9chh4 lactococcus
37	154	7.0	1283	3	Q9USQ3	Q9usq3 schizosacch
38	154	7.0	1283	3	Q9URU4	Q9uru4 schizosacch
39	153.5	7.0	1795	5	O76894	O76894 drosophila
40	153.5	7.0	3570	4	Q99552	Q99552 homo sapien
41	153	7.0	338	4	O43420	O43420 homo sapien
42	153	7.0	769	5	Q17921	Q17921 caenorhabdi
43	151	6.9	961	3	Q92223	Q92223 emericella
44	151	6.9	2586	5	Q9VTK8	Q9vtk8 drosophila
45	149	6.8	966	5	Q9U3P1	Q9u3p1 caenorhabdi

ALIGNMENTS

```
RESULT   1
Q9Y743
ID   Q9Y743      PRELIMINARY;      PRT;  1270 AA.
AC   Q9Y743;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN.
GN   ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=21064501; PubMed=11124701;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS5 gene of Candida albicans and analysis of the Als5p N-
RT   terminal domain.";
RL   Yeast 18:49-60(2001).
DR   EMBL; AF068866; AAD32849.1; -.
SQ   SEQUENCE   1270 AA;  133337 MW;  D8E5FA5853F6D5C5 CRC64;
```

```
  Query Match             87.8%;  Score 1922;  DB 3;  Length 1270;
  Best Local Similarity   87.0%;  Pred. No. 1.7e-116;
  Matches  362;  Conservative   20;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || |||:|:| :||||||| ||||||||||||||||||||||||||||||| ||||||||
Db      17 AKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTSANPGDTFILNMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           :| || |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 FTASQKSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           |||||||||||||||| |||||||| | :|| | ||||||| | ||  || ||||||:
```

```
Db    137 GGTGSSTDLEDSKCFTAGINTVTFNDGSKKLSIAVNFEKSTVDRSGYLTTSRFMPSLNKI 196

Qy    181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240

Db    197 ATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSC 256

Qy    241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300

Db    257 SSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYK 316

Qy    301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360

Db    317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy    361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416

Db    377 SYSTKTAPIGETATLIVDVPYHTTTTVTSKWTGTITMTTTRTNPTDSIDTVVVQVP 432


RESULT   2
O93865
ID   O93865      PRELIMINARY;      PRT;  1047 AA.
AC   O93865;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE CELL SURFACE PROTEIN.
GN   ALS8.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC10261;
RA   Leng P., Lee P.R., Wishart J.A., Wu H., Brown A.J.P.;
RT   "Sequence of the hypha-specific, agglutinin-like cell surface protein,
RT   ALS8 from Candida albicans.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF051313; AAD02580.1; -.
SQ   SEQUENCE   1047 AA;  111944 MW;  C2327659AA911F2E CRC64;


  Query Match             86.3%;  Score 1889;  DB 3;  Length 1047;
  Best Local Similarity   84.9%;  Pred. No. 1.9e-114;
  Matches  353;  Conservative   30;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60

Db     17 AKTITGVFNSFNSLTWSNAATYHYKGPGTPTWNAVLGWSLDGTSASPGDTFTLNMPCVFK 76

Qy     61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120

Db     77 FTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNV 136

Qy    121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180

Db    137 GGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKV 196

Qy    181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240

Db    197 STLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTC 256

Qy    241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300

Db    257 SSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYR 316

Qy    301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360

Db    317 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTITTSYVGVTT 376

Qy    361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416

Db    377 SYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVP 432


RESULT   3
Q9URQ0
ID   Q9URQ0      PRELIMINARY;      PRT;  468 AA.
AC   Q9URQ0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
DR   EMBL; AF024582; AAC64237.1; -.
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50143 MW;  2E38E36D7FEAD192 CRC64;


  Query Match             80.6%;  Score 1764.5;  DB 3;  Length 468;
  Best Local Similarity   79.8%;  Pred. No. 8.7e-107;
  Matches  332;  Conservative   28;  Mismatches   55;  Indels    1;  Gaps    1;

Qy      1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60

Db     17 AKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76

Qy     61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120

Db     77 FITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTSSIKALGTVTLPISFNV 136

Qy    121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180

Db    137 GGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDASGFFIASRLIPSINKV 196

Qy    181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240

Db    197 SITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLNDWNFPVSSDSLSYNKTC 256

Qy    241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300

Db    257 SSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVGSSLQSKPFNLRLRGYN 315

Qy    301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360

Db    316 NSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 375

Qy    361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416

Db    376 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITATTTRTNPTDSIDTVVVQVP 431


RESULT   4
Q9HFX4
ID   Q9HFX4      PRELIMINARY;      PRT;  1523 AA.
AC   Q9HFX4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
```

```
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SC5314;
RA   Chen X., Chen J.-Y.;
RT   "ALS4 (agglutinin-like sequence) of Candida albicans.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF272027; AAG25054.1; -.
FT   NON_TER    1523    1523
SQ   SEQUENCE   1523 AA;  159168 MW;  73AF3B3E442FD53C CRC64;


  Query Match             73.3%;  Score 1605;  DB 3;  Length 1523;
  Best Local Similarity   70.9%;  Pred. No. 7.3e-96;
  Matches  295;  Conservative   45;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || |||:|||||||||:|||:|:::||  ||| ||:|||||| :|: ||||||:||||||
Db      17 AKVITGIFDSFNSLTWTNAASYSYRGPANPTWTAVIGWSLDGATASAGDTFTLDMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : | |||:|| |||  ||||   | |||||||:::|||   : :  || |||||| :|:|
Db      77 FITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTADTKAIGTVTLPFSFSV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           ||:||  || :|:||||| |||||||||  ||  |:|||||  |  :  ||::|||::
Db     137 GGSGSDVDLANSQCFTAGINTVTFNDGDTSISATVDFEKSTVASSDRILLSRILPSLSQA 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
            :||:  :| ||||||||||::     |||| :|:||: ||||||||:||||||||||
Db     197 VSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           ||  : : ||||||||||:|||:||| |: | : ||| | | |:      ||  |:||
Db     257 TSTSVLVTYQNVPAGYRPFVDAYVSATRYSSYAMRYTNIYACVGAASVDDSFTHTWSGYS 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           || |||||| || ||||||||||||||||   |||||||||||||||||||||||||||||
Db     317 NSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSESDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432


RESULT   5
Q9URP8
ID   Q9URP8      PRELIMINARY;      PRT;   469 AA.
AC   Q9URP8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
DR   EMBL; AF024586; AAC64241.1; -.
FT   NON_TER     469     469
SQ   SEQUENCE   469 AA;  49597 MW;  88BC96D79142C8DB CRC64;


  Query Match             73.2%;  Score 1604;  DB 3;  Length 469;
  Best Local Similarity   71.2%;  Pred. No. 2.2e-96;
  Matches  296;  Conservative   43;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           || |||||:|||||||:|||:| ::||  ||| ||:|||||| :|: ||||||:||||||
Db      17 AKVITGVFNSFNSLTWANAASYPYRGPATPTWTAVIGWSLDGATASAGDTFTLDMPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : | |||:|| |||  ||||   | |||||||:::|||   : :  || |||||| :|:|
Db      77 FITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTADTKAMGTVTLPFSFSV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           ||:||  || :|:||||| |||||||||  ||  |:|||||  |  :  ||::|||::
Db     137 GGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVASSDRILLSRILPSLSQA 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
            :||:  :| ||||||||||::     |||| :|:||: ||||||||:||||||||||
Db     197 VSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTC 256

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
           ||  : : ||||||||||:|||:||| |: | : ||| | | |:      ||  |:||
Db     257 TSTSVLVTYQNVPAGYRPFVDAYVSATRVSSYTMQYTNIYACVGAASVDDSFTHTWRGYS 316

Qy     301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
           || |||||| || ||||||||||||||||   |||||||||||||||||||||||||||||
Db     317 NSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSDTDKTKTIEILQPIPTTTITTSYVGVTT 376

Qy     361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 SYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 432


RESULT   6
Q9C471
ID   Q9C471      PRELIMINARY;      PRT;   468 AA.
AC   Q9C471;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN   ALS9.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RA   Hoyer L.L., Hecht J.E., Mirus K.A.;
RT   "The ALS9 gene of Candida albicans.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229989; AAK00764.1; -.
FT   NON_TER     468     468
SQ   SEQUENCE   468 AA;  50127 MW;  B291D3EB15FB96DE CRC64;


  Query Match             73.2%;  Score 1602.5;  DB 3;  Length 468;
  Best Local Similarity   71.6%;  Pred. No. 2.7e-96;
  Matches  298;  Conservative   48;  Mismatches   69;  Indels    1;  Gaps    1;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           ||||||||:||:||||: :  ||:|||  ||||||||||: |:|:|||||||| :|||||
Db      17 AKTITGVFNSFDSLTWTRSVEYAYKGPETPTWNAVLGWSLNSTTADPGDTFTLILPCVFK 76

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : |:||||||||| |||| | :|||||||:|:||||     |  :  ||| ||| |||
Db      77 FITTQTSVDLTADGVSYATCDFNAGEEFTTFSSLSCTVNSVSVSYARVSGTVKLPITFNV 136

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           |||||| || |||||||| ||||| |||  ||  |:|: | | || |: :||::|||||:
Db     137 GGTGSSVDLADSKCFTAGKNTVTFMDGDTKISTTVDFDASPVSPSGYITSSRIIPSLNKL 196

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           ::||| |||||||||| ||| :|||   |||||::|||:|||||||:||||||||||||
```

```
Db    197 SSLFVVPQCENGYTSGIMGFVASNG-ATIDCSNVNIGISKGLNDWNFPVSSESFSYTKTC 255

Qy    241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYK 300
          ||  | :::||||||||||:|||||| ::::||| | |:||          |||| | |||
Db    256 TSTSITVEFQNVPAGYRPFVDAYISAENIDKYTLTYANEYTCENGNTVVDPFTLTWWGYK 315

Qy    301 NSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTT 360
          ||:| |:| ||| |||||||||||||||||||||:||||||||||||||||:|::|
Db    316 NSEADSDGDVIVVTTRTVTDSTTAVTTLPFNPSVDKTETIEILQPIPTTTITTSYIGIST 375

Qy    361 SYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
          || | |  || ||||||| ||| | |||: |||:|||||| |||| |||||:||:|
Db    376 SYETLTGTIGGTATVIVDTPYHITATVTNFWTGSITTTTTYTNPTGSIDTVIVQIP 431


RESULT   7
Q9Y8F2
ID   Q9Y8F2      PRELIMINARY;      PRT;  1443 AA.
AC   Q9Y8F2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN 6.
GN   ALS6.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF075293; AAD42033.1; -.
DR   InterPro; IPR003015; HLH_Myc.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ   SEQUENCE   1443 AA;  151476 MW;  DCBD693F4C435809 CRC64;


  Query Match             70.7%;  Score 1549;  DB 3;  Length 1443;
  Best Local Similarity   68.7%;  Pred. No. 2.9e-92;
  Matches  287;  Conservative   54;  Mismatches   75;  Indels    2;  Gaps    1;

Qy      1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
          ||||:||| ||||||:::|  || : |||||| |||||||| |:|||||| ||||||
Db     18 AKTISGVFTSFNSLTYTNTGNYPYGGPGYPTWTAVLGWSLDGTLASPGDTFTLVMPCVFK 77

Qy     61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
          : |:||||||||:|||||| |::||:||||:::| ||: | |:|:|||| |||:|||
Db     78 FITTQTSVDLTANGVKYATCTFHAGEDFTTFSSMSCVVNNGLSSNIRAFGTVRLPISFNV 137

Qy    121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
          |||||| :::|||||||||||||| |||  ||  | | ||:    |: :| :||:|||:|:
Db    138 GGTGSSVNIQDSKCFTAGTNTVTFTDGDHKISTTVNFPKTPQSSSSLVYFARVIPSLDKL 197

Qy    181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
          ::| || ||  || || :|||::  || |||| ||:||| ||| || ||||||||||||
Db    198 SSLVVASQCTAGYASGVLGFSATKDDVTIDCSTIHVGITNGLNSWNMPVSSESFSYTKTC 257

Qy    241 TSNGIQIKYQNVPAGYRPFIDAYI--SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTG 298
          | |   | |:||||||||||:|:  |||  | : | ||| | |   :  : |    ||
Db    258 TPNSFIITYENVPAGYRPFIDSYVKKSATATNGFNLNYTNIYNCMDGKKGNDPLIYFWTS 317

Qy    299 YKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGV 358
          | ||||||||  :| ||||||||||:|||||:|:|||||||:::|||||||||||||:
Db    318 YTNSDAGSNGAAVVVTTRTVTDSTTAITTLPFDPTVDKTKTIEVIEPIPTTTITTSYVGI 377

Qy    359 TTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
          :||  |||| || ||||:||||||||||:|| |||: ||::| ||||||||||||||
Db    378 STSLSTKTATIGGTATVVVDVPYHTTTTITSIWTGSATTSSTYTNPTDSIDTVVVQVP 435


RESULT   8
Q9HF70
ID   Q9HF70      PRELIMINARY;      PRT;  336 AA.
AC   Q9HF70;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALSD2P (FRAGMENT).
GN   ALSD2.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202529; AAG35623.2; -.
FT   NON_TER       1      1
FT   NON_TER     336    336
SQ   SEQUENCE   336 AA;  35593 MW;  E513E6EA9E8E9EC7 CRC64;


  Query Match             67.2%;  Score 1471;  DB 3;  Length 336;
  Best Local Similarity   81.0%;  Pred. No. 6.2e-88;
  Matches  272;  Conservative   31;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     45 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 104
          |:||||||| |||||: |:|||||||:||||||| ||||||||||:|||||| || |
Db      1 ASPGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCSFYSGEEFTTFSSLTCTVNSALTS 60

Qy    105 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 164
          |:|||||||||:|||||||| |||||||||||||||| |||  :|  |:|||||||
Db     61 SVKAFGTVTLPISFNVGGTGSSVDLEDSKCFTAGTNTVTFTDGDNKVSTTVDFEKSTVDS 120

Qy    165 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 224
          : || :||:|||||||:|||||||  ||||||:||||||| |: ||||:|:||| |:||
Db    121 TGYLTSSRLMPSLNKVTSLFVAPQCARGYTSGTIGFSSSNGGVSFDCSNVHVGITNGVND 180

Qy    225 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 284
          ||:|||||||||||:| || | |:|||||||||||||:|:||| |||:||||||| |
Db    181 WNFPVSSESFSYTKTCSSTGITITYRNVPAGYRPFIDAYITASDVNSYTLSYTNDYTCVG 240

Qy    285 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 344
            :| |||||||:|||| :|||||||||||||||||||||||||||  ||||||||||
Db    241 GSVQHKPFTLRWSGYKNGEAGSNGIVIVATTRTVTDSTTAVTTLPFNSDVDKTKTIEILQ 300

Qy    345 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 380
          |||||||| |||||||| |||||||:|||: :|:|
Db    301 PIPTTTITTLYVGVTTSYSTKTAPIGDTATLFIDIP 336


RESULT   9
Q9HF72
ID   Q9HF72      PRELIMINARY;      PRT;  338 AA.
AC   Q9HF72;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALSD1P (FRAGMENT).
GN   ALSD1.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
```

```
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201685; AAG35603.1; -.
FT   NON_TER       1      1
FT   NON_TER     338    338
SQ   SEQUENCE   338 AA;  35908 MW;  C0A7F2F94609E172 CRC64;


  Query Match             54.4%;  Score 1191;  DB 3;  Length 338;
  Best Local Similarity   64.8%;  Pred. No. 8.7e-70;
  Matches  219;  Conservative   51;  Mismatches   66;  Indels    2;  Gaps    1;

Qy      45 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 104
           || |||||| |||||: |:|||||||:|||||| |::||:|| ||:::| ||: | |
Db       1 ANAGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCTFHAGEDFTAFSSMSCVVNNGLTS 60

Qy     105 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 164
           :||||||| :||:|||||||||| :|:|||||||||||:||| |||  ||| |:| |:
Db      61 NIKAFGTVRIPISFNVGGTGSSVNLQDSKCFTAGTNSVTFTDGDHKISIPVDFPKTPESS 120

Qy     165 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 224
           |  :  |||:|:|:|:::| || ||  || || :|||::  || |:||||:|:||| |||
Db     121 SGLIKYSRVIPTLDKLSSLAVASQCTAGYKSGVLGFSATKNDVTIECSNVHVGITNGLNS 180

Qy     225 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQ--YTLAYTNDYTC 282
           || ||||:|||||||||:    | |:|||||||||| |:   |       : | ||| | |
Db     181 WNMPVSSDSFSYTKTCTSSSFIITYENVPAGYRPFIDTYVKKTSTTSTGFNLNYTNSYVC 240

Qy     283 AGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 342
           :  : | |   || | ||||||:| |:: ||:||||||||:|||||:|:||||||||:
Db     241 TDGKKGNDPLIYFWTSYTNSDAGSDGAVVIVTTKTVTDSTTAITTLPFDPTVDKTKTIEV 300

Qy     343 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 380
           |:|||||||||||||:||: |||| |||||:|:|||
Db     301 LEPIPTTTITTSYVGVSTSPFTKTATIGETATLIIDVP 338


RESULT  10
Q9HF69
ID   Q9HF69      PRELIMINARY;      PRT;   331 AA.
AC   Q9HF69;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALSD3P (FRAGMENT).
GN   ALSD3.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202530; AAG35624.2; -.
FT   NON_TER       1      1
FT   NON_TER     331    331
SQ   SEQUENCE   331 AA;  35297 MW;  469AB72F9CE029BF CRC64;


  Query Match             53.7%;  Score 1176;  DB 3;  Length 331;
  Best Local Similarity   64.3%;  Pred. No. 8e-69;
  Matches  214;  Conservative   44;  Mismatches   73;  Indels    2;  Gaps    1;

Qy      45 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 104
           || |||||||||||||: | |||||| |||  ||||  |||||||||:| |||::|| |
Db       1 ANAGDTFTLNMPCVFKFITDQTSVDLVADGRTYATCDLYSGEEFTTFSSLKCTVSNALNS 60

Qy     105 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 164
           || |||||::||:||:|| |:  |:||  |||||||||||  |    |::| |:
Db      61 QTKALGTVTLPLSFNIGGSGSDVDITSSQCFKEGTNTVTFNDGDTTFSTTANFQRSDVNA 120

Qy     165 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 224
           :  :  ||::|||  |  |:|: |:| :|||:|||||||::  |  |||| :| ||: ||||
Db     121 NDRILLSRILPSLAKSVTIFIPPRCASGYSSGTMGFSTAGTDAIIDCSTVHAGISNGLND 180

Qy     225 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 284
           ||||: |:|||||  :|:| |: : ||||||||||||:||||||  :  ||: ||| ||| |
Db     181 WNYPIESKSFSYTTSCSSKGVSVTYQNVPAGYRPFVDAYISA--LTSYTMQYTNQYTCVG 238

Qy     285 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 344
           :|      |:  | || |::||| || || || ||||||||||||||  ||:|||| :||
Db     239 ARPVDASFSYNWLGYDNAEAGSRGITIVVTTSTVTDSTTAVTTLPFNSDVDRTKTIAVLQ 298

Qy     345 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIV 377
           |||||||||||||||||| |:| |||||||: :
Db     299 PIPTTTITTSYVGVTTSYRTQTVPIGETATLFI 331


RESULT  11
Q9HGK6
ID   Q9HGK6      PRELIMINARY;      PRT;  2297 AA.
AC   Q9HGK6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALS7P.
GN   ALS7.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF201684; AAF98068.1; -.
SQ   SEQUENCE   2297 AA;  244723 MW;  59B020C63027F651 CRC64;


  Query Match             45.5%;  Score 997.5;  DB 3;  Length 2297;
  Best Local Similarity   47.0%;  Pred. No. 2.8e-56;
  Matches  196;  Conservative   71;  Mismatches  149;  Indels    1;  Gaps    1;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           :| :||||: |||| ||      ::     | || | |:|||| |:||||||| ||||:|
Db      18 SKEVTGVFNQFNSLIWSYTYRARYEEISTLTANAQLEWALDGTIASPGDTFTLVMPCVYK 77

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : | :||| |||: : |||| | :||:  :||:| ||| | |      ||:| |||||||
Db      78 FMTYETSVQLTANSIAYATCDFDAGEDTKSFSSLKCTVTDELTEDTSVFGSVILPIAFNV 137

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           ||:|| : : ||||||:| ||||| ||:  :|    |          : : |:  ||: :
Db     138 GGSGSKSTITDSKCFSSGYNTVTFFDGNNQLSTTANFLPRRELAFGLVVSQRLSMSLDTM 197

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           |   ::  |  || || :||:|:: |  ||||:||:|||  :|||: |||| | :|  |
Db     198 TNFVMSTPCFMGYQSGKLGFTSNDDDFEIDCSSIHVGITNEINDWSMPVSSVPFDHTIRC 257

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWT-GY 299
           ||  : |::: :||||||:|| :        : : |||:: |       | |||  ::
Db     258 TSRALYIEFKTIPAGYRPFVDAIVQIPTTEPFFVKYTNEFACVNGIYTSIPFTSFFSQPI 317

Qy     300 KNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVT 359
           :| : |  :| || ||  | |  |||||   : |||||:|:|||||:|||: |
Db     318 LYDEALAIGADLVRTTSTVIGSITRTTTLPFISRLQKTKTILVLEPIPTTTVTTSHHGFD 377

Qy     360 TSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVP 416
```

```
          | | || | ||:|||| :||| || ||:|: |  : | |||  :  | :|||:|::|
Db    378 TWYYTKKATIGDTATVFIDVPQHTATTLTTYWQESSTATTTYFDDIDLVDTVIVKIP 434


RESULT  12
Q9Y8F1
ID   Q9Y8F1      PRELIMINARY;      PRT;   433 AA.
AC   Q9Y8F1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN 7 (FRAGMENT).
GN   ALS7.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF075294; AAD42034.1; -.
FT   NON_TER     433     433
SQ   SEQUENCE   433 AA;  48037 MW;  1F30CFD99C2EC445 CRC64;


  Query Match             45.2%;  Score 990.5;  DB 3;  Length 433;
  Best Local Similarity   46.9%;  Pred. No. 1.1e-56;
  Matches  195;  Conservative   71;  Mismatches  149;  Indels    1;  Gaps    1;

Qy       1 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 60
           :| :||||: |||| ||      ::     | || | |:|||| |:||||||| ||||:|
Db      18 SKEVTGVFNQFNSLIWSYTYRARYEEISTLTANAQLEWALDGTIASPGDTFTLVMPCVYK 77

Qy      61 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
           : | :||| |||: : |||| | :||:  :||:| ||| | |       ||:| |||||||
Db      78 FMTYETSVQLTANSIAYATCDFDAGEDTKSFSSLKCTVTDELTEDTSVFGSVILPIAFNV 137

Qy     121 GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
           ||:|| : : |||||::| ||||| ||:  :|    |          : : |:  ||: :
Db     138 GGSGSKSTITDSKCFSSGYNTVTFFDGNNQLSTTANFLPRRELAFGLVVSQRLSMSLDTM 197

Qy     181 TTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC 240
           |   ::  |  || || :||:|:: |  ||||:||:||| :|||: |||| | :|  |
Db     198 TNFVMSTPCFMGYQSGKLGFTSNDDDFEIDCSSIHVGITNEINDWSMPVSSVPFDHTIRC 257

Qy     241 TSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWT-GY 299
           ||  : |::: :|||||||:|| :      : : |||:: |      | |||  ::
Db     258 TSRALYIEFKTIPAGYRPFVDAIVQIPTTEPPFVKYTNEFACVNGIYTSIPFTSFFSQPI 317

Qy     300 KNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVT 359
             :| : |  :| || ||  | |  |||||    : ||||| :|:||||||:|||: |
Db     318 LYDEALAIGADLVRTTSTVIGSITRTTTLPFISRLQKTKTILVLEPIPTTTVTTSHHGFD 377

Qy     360 TSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQV 415
           | | || | ||:|||| :||| || ||:|: |  : | |||  :  | :|||:|::
Db     378 TWYYTKKATIGDTATVFIDVPQHTATTLTTYWQESSTATTTYFDDIDLVDTVIVKI 433


RESULT  13
Q9HF71
ID   Q9HF71      PRELIMINARY;      PRT;   353 AA.
AC   Q9HF71;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN 1 (FRAGMENT).
GN   ALST1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13803;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201686; AAG35604.1; -.
FT   NON_TER       1       1
FT   NON_TER     353     353
SQ   SEQUENCE   353 AA;  37272 MW;  4268154D5615DE08 CRC64;


  Query Match             44.8%;  Score 981;  DB 3;  Length 353;
  Best Local Similarity   54.9%;  Pred. No. 3.7e-56;
  Matches  185;  Conservative   50;  Mismatches  100;  Indels    2;  Gaps    2;

Qy      44 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 103
           :|| |||||| ||||||:|||:||:|||      |||| | :|| |||||:|:|||  ::
Db      19 AANAGDTFTLIMPCVFKFTTSETSIDLTVGSKSYATCNFNAGEHFTTFSSLSCTVTQSVP 78

Qy     104 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 163
             :  |:||:|:|:|||||||:|   || |:|||| | |||||:|||  |    || :
Db      79 DNTNAYGTITVPLAFNVGGSGRDVDLTDAKCFTTGDNTVTFSDGDKSFSTTANFEGAGTL 138

Qy     164 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 223
             |  :||::||| |   | ||| | ||| |||:|||||:    :|||:||  |||  ||
Db     139 NDDY-ESSRLIPSLGKTDALLVAPLCSNGYKSGTIGFSSTTKGFSIDCNNIQAGITSQLN 197

Qy     224 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 283
             | :|  |:|||||  ||:    | :  :| | |||||||| |   : | : ||  | |:
Db     198 AWGFPTDSQSFSYTTQCTTTSYSITFSTIPKGLRPFIDAYIKA-PTSTYPMTYTYKYVCS 256

Qy     284 GSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEIL 343
             :  :    | |:|| |||| | |: ||  | | | ||| ||||||: : ||||||:::
Db     257 DGKSYNGNTKLNWSGYVNSDADSEGMEIVVATTTGTGSTTGVTTLPFDKTKDKTKTIQVI 316

Qy     344 QPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 380
           :||||||:||||:||||||: | || ||||||:::|:|
Db     317 EPIPTTTVTTSYLGVTTSFSTITATIGETATLVIDMP 353


RESULT  14
Q9HF49
ID   Q9HF49      PRELIMINARY;      PRT;   126 AA.
AC   Q9HF49;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALST3P (FRAGMENT).
GN   ALST3.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13803;
RX   MEDLINE=21186039; PubMed=11290712;
RA   Hoyer L.L., Fundyga R., Hecht J.E., Kapteyn J.C., Klis F.M.,
RA   Arnold J.;
RT   "Characterization of agglutinin-like sequence genes from non-albicans
RT   candida and phylogenetic analysis of the ALS family.";
RL   Genetics 157:1555-1567(2001).
DR   EMBL; AF211866; AAG43533.1; -.
FT   NON_TER       1       1
FT   NON_TER     126     126
SQ   SEQUENCE   126 AA;  13419 MW;  D28914A425526888 CRC64;


  Query Match             16.4%;  Score 359.5;  DB 3;  Length 126;
```

```
  Best Local Similarity   55.1%;  Pred. No. 2.1e-16;
  Matches   70;  Conservative   15;  Mismatches   41;  Indels    1;  Gaps    1;

Qy     254 AGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVA 313
           ||||||  | : |   : | : ||  | | ||  :     : | || |||  ||| |:|
Db       1 AGYRPFFAALVQAPS-SDYAIQYTAKYRCEGSVQRDDSQKISWAGYTNSDPDSNGAVVVL 59

Qy     314 TTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETA 373
           || | | | | ||||||||: | |||||:: |||| | ||||:|||||| | |  ||:||
Db      60 TTITGTQSNTIVTTLPFNPTADHTKTIEVIVPIPTVTTTTSYIGVTTSYTTITGTIGDTA 119

Qy     374 TVIVDVP 380
           |:::|:|
Db     120 TLVIDMP 126


RESULT  15
Q9HF50
ID   Q9HF50      PRELIMINARY;      PRT;   127 AA.
AC   Q9HF50;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AGGLUTININ-LIKE PROTEIN ALST2P (FRAGMENT).
GN   ALST2.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13803;
RX   MEDLINE=21186039; PubMed=11290712;
RA   Hoyer L.L., Fundyga R., Hecht J.E., Kapteyn J.C., Klis F.M.,
RA   Arnold J.;
RT   "Characterization of agglutinin-like sequence genes from non-albicans
RT   candida and phylogenetic analysis of the ALS family.";
RL   Genetics 157:1555-1567(2001).
DR   EMBL; AF211865; AAG43532.1; -.
FT   NON_TER       1      1
FT   NON_TER     127    127
SQ   SEQUENCE   127 AA;  13668 MW;  178E25AA919EF387 CRC64;


  Query Match            15.6%;  Score 341;  DB 3;  Length 127;
  Best Local Similarity   51.2%;  Pred. No. 3.4e-15;
  Matches   65;  Conservative   22;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     254 AGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVA 313
           ||||||:|   | |  : :|: |||:|  |            |  |::|    :| :|:
Db       1 AGYRPFVDVLFSHTASDIFTMLYTNEYVGADGVTYDASMKKAWKSYQDSLPSGDGAIIIV 60

Qy     314 TTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETA 373
           |||| | |||||:|||::| :| |||||:| |||||| ||||:||:| | | || ||:||
Db      61 TTRTGTQSTTAVSTLPYDPEIDLTKTIEVLVPIPTTTTTTSYLGVSTYYSTITATIGDTA 120

Qy     374 TVIVDVP 380
           |:::|:|
Db     121 TLVIDMP 127


Search completed: October  3, 2002, 16:33:57
Job time: 250 sec
```